Sensing and Visualizing Social Context from Spatial Proximity

Thesis

submitted in partial fulfillment of the requirements of the degree of
Doktor-Ingenieur (Dr.-Ing) of the Faculty of Mathematics and Informatics
University of Bremen, Bremen, Germany

by

Tom Nicolai

Bremen, Dec. 7th, 2010

Advisor: Prof. Dr. Otthein Herzog (University of Bremen, Germany)
Second advisor: Prof. Dr. Jonathan Crowcroft (University of Cambridge, UK)

Disputation date: Dec. 2nd, 2011

Abstract

The concept of pervasive computing, as introduced by Marc Weiser under the name ubiquitous computing in the early 90s, spurred research into various kinds of context-aware systems and applications. There is a wide range of contextual parameters, including location, time, temperature, devices and people in proximity, which have been part of the initial ideas about context-aware computing. While locational context is already a well understood concept, social context—based on the people around us—proves to be harder to grasp and to operationalize.

This work continues the line of research into social context, which is based on the proximity and meeting patterns of people in the physical space. It takes this research out of the lab and out of well controlled situations into our urban environments, which are full of ambiguity and opportunities.

The key to this research is the tool that caused dramatic change in individual and collective behavior during the last 20 years and which is a manifestation of many of the ideas of the pervasive computing paradigm: the mobile phone. In this work, the mobile is regarded as a proxy for people. Through it, the social environment becomes accessible to digital measurement and processing. To understand the large amount of data that now becomes available to automatic measurement, we will turn to the discipline of social network analysis. It provides powerful methods, that are able to condense data and extract relevant meaning. Visualization helps to understand and interpret the results.

This thesis contains a number of experiments, that demonstrate how the automatic measurement of social proximity data through Bluetooth can be used to measure variables of personal behavior, group behavior and the behavior of groups in relation to places. The principal contributions are:

- A methodology to visualize personal social context by using an ego proximity network. Specific episodes can be localized and compared.
- A method to compare different days in terms of social context, e.g. to support automatic diary applications.
- A method to compose social geographic maps. Locations of similar social context are detected and combined.

- Functions to measure short-term changes in social activity, based on the distinction between strange and familiar devices.
- The characterization of Bluetooth inquiries for social proximity sensing.
- A dataset of Bluetooth sightings from an ego perspective in seven different settings. Additionally, some settings feature multiple stationary scanners and Cell-ID measurements.
- Soft- and hardware to capture, collect, store and analyze Bluetooth proximity data.

Acknowledgments

This work would not have been possible without the great support of a number of people. First, I would like to express my sincere gratitude to my supervisor at the TZI in Bremen Prof. Otthein Herzog, who gave me the freedom to pursue this research. Prof. Jon Crowcroft from the University of Cambridge has been exceptionally supportive in giving my thesis the final focus.

At the TZI, all the members of the vibrant wearable computing group had an impact on the outcome of this work. Especially, Dr. Ingrid Rügge with unconditional trust in my work (and a skeptical view at the same time), Dr. Holger Kenn with his never-ending interest in everything new and Dr. Michael Boronowsky, the visionary behind the wearable computing idea in Bremen, have supported my work in particular.

Regarding the topic of Bluetooth scanning, the cooperation with Prof. Vassilis Kostakos at the University of Bath and the Cityware project have been a true enlightenment. The joint experiments with Dr. Eiko Yoneki from the University of Cambridge during the PerCom '06 and Ubicomp '06 have been a pleasure and substantiated my work with valuable data.

I would also like to thank Dr. Hendrik Witt for a lot of controversial discussions, Nils Behrens for his indispensable work on the WirelessRope software and support with the PerCom06 experiment, Andreas Kemnade for his fabulous work on the TrackStations programs and hardware, Rüdiger Leibrandt for building and soldering the hardware of the TrackStations, Xin Xing for her work on the WirelessRope software, Autostart and Watchdog programs, Olena Rohovets for her support with augmented gatecounts in Bremen and additional work on the TrackStation software, Alena Braun for coding on the ropeviz program and Frauke Buisman for helping me conduct more augmented gatecounts in Bremen.

Probably, I could not have completed this work without the encouragement and personal dedication of Charlotte Irmler.

Contents

List of Figures

List of Tables

Part I.

Starting Situation and Related Work

1. Introduction

Mario Tokoro, President of the Sony CS Lab, gave the keynote speech at Ubicomp05. During the usual beginning with a time line of computing history ("the computer was invented in ..."), among all the common items of hardware and operating systems, I noted the small words "Blog," "Wiki" and "SNS" and they caught my curiosity.

Within his speech, he distinguished three major paradigms in computing. First, computer programs were designed as tools: spreadsheets, word processors, etc. The next thing was the pervasive computing idea. Computers integrate into the environment, into everyday things, into clothes. The personal computer disappears.

As the third paradigm he brought up an issue that came as a surprise. He stated, that computers will be used "as a means to build society." He mentioned Blogs, Wikis and social networking services as indications. He also talked about new evolving sciences and the importance of interdisciplinary research. I really enjoyed his keynote.

With my thesis in hand, I try to follow the visionary direction, that Mario Tokoro outlined in his keynote speech. In doing so, I will draw connections between seemingly unrelated things. As a computer scientist, I will borrow from the social sciences. Many hypotheses are based on simplified assumptions about the behavior of people. And after proving my method of measurement to be unreliable, I will use it anyway. My approach may seem naive at some points—but the results do not flow without any logic.

1.1. Background: Anonymity and the City

With the formation of cities during industrialization, a new style of living developed. Although people lived closer together spatially than before in rural areas, they were gaining more privacy [64]. This increased privacy was mainly achieved through the anonymity cities provide in comparison to rural villages. Nowadays, citizens hardly know everybody living in the same building, let alone in their neighborhood. We enjoy this urban anonymity, but we also feel alone and have no one to ask for a helping hand. Thus, urbanization has made cooperation more difficult, but at the same time, has increased the opportunities for cooperation tremendously. We meet hundreds of people while taking an urban walk, but whom can we ask for a specific favor? Who would be willing and who could be able to provide help? On a larger scale, cooperation between individuals would lead to efficient sharing of resources and thus to a more effective operation of the city as a whole. Pervasive computing has the potential to uncover these hidden opportunities in a natural and spontaneous way, but for urban applications to function effectively, there is the need to precisely measure, describe and analyze the infrastructure that a specific city provides.

Although this topic is discussed controversially, providing a local community with high speed Internet access and basic tools for communication helps to increase the resident's social networks, i.e. they get in contact with more people or increase the intensity of existing contacts. A study about such a networked village has shown that the residents especially formed weak ties with others and used them for the purpose of collective action [63]. Earlier investigations of Internet usage in general showed complex effects on the social behavior of users [97, 138, 137], both stimulating and weakening, depending on the personality of people.

As the field of wireless and locative technologies matures, a more enduring relationship between the physical and cultural elements and its digital topographies are becoming interesting topics to explore. Their interaction, influence, disruption, expansion and integration with the social and material practices of our public spaces are receiving more focus. Is public space a crowd of individuals? How can the crowd inspire the individual through collaboration, competition and confrontation? How could change, effect or experience be achieved by a mass movement, a cooperative crowd? How could we stage a series of new happenings?

1.2. Social Context in Pervasive Computing

When we consider the evolution of the three paradigms outlined by Mario Tokoro, we should be able to find the roots for the social perspective in pervasive computing. Let us take a short tour into this exciting field of research.

In the influential article "The Computer for the 21st Century," Mark Weiser [196] introduced the concept of pervasive computing (originally called ubiquitous computing), shifting the focus of computing research from the *general purpose computer* towards computers that are more *tightly integrated into the working processes and social lives* of the users. Lamming and Flynn [103] created a prototypical mobile computer with the goal to support human memory by automatically recording meetings with other persons as well as interactions. Real-life meetings between the users of the devices were detected by short-range radio. Whenever a device detected another one, this event was recorded in memory. The current location of users could be measured with the same technique, employing stationary devices in the environment as identifiers for specific locations. Users were also able to store documents on their devices. The short-range radio was used to exchange these documents during real-life meetings. Through the integration of proximity information and interaction, a diary was automatically created in the device. Thus, Lamming and Flynn found that it was easy for the users to keep track of their activities. They could easily look up, when they met a specific person and what the meeting was about.

Kortuem and Segall [92] extended these ideas by introducing the concept of *wearable communities* in contrast to *virtual communities*. The goal of their research was to augment real-life communities and enable mutual cooperation between the members. They explored several applications, including *Genie* and *WALID*. With Genie, users could store questions in their devices that were automatically transmitted and presented to nearby users of the same devices. The questions served the purpose to identify and get into contact with persons sharing the same interests. WALID, on the other hand, was based on shared task-lists. Every user could enter tasks that were matched with nearby users. When an overlap in the lists was detected, the users could decide to help each other for mutual benefit.

Recently, the *Human Dynamics Group* at the MIT picked up the work on proximity sensing to automatically measure social networks [45, 57]. Social network analysis is a tool usually applied by sociologists to analyze various social settings, including communication structures in companies or family structures. In several large-scale experiments, they showed how detailed information about social networks could be gathered by infrared, audio and Bluetooth technologies. This data even proved to reveal different relationships to others, including friends and colleagues.

While all these works focus on real-life, spatial proximity to people whose identity is known—if not to the person, then at least to the system—Paulos' and Goodmans' research concentrated on proximity to strangers in public places [149]. They used Bluetooth technology present in many modern mobile phones to measure characteristics about the social situation of individuals. Their device could measure, if the user was in a familiar or strange social setting. O'Neill et al. gathered and analyzed similar data, that was acquired by stationary Bluetooth scanners throughout the city of Bath [145]. They demonstrated, that it was possible to measure movement in public places and to build social networks from the data.

1.3. Contributions

This work continues the line of research into social context, which is based on the proximity and meeting patterns of people in the physical space. It takes this research out of the lab and out of well controlled situations into urban environments, which are full of ambiguity and opportunities. To characterize these complex environments, a concept, called the *urban pervasive infrastructure*, is introduced. It provides a framework, in which we will focus on the proximity patterns measured by Bluetooth device inquiries and Cell-ID readings.

The key to this research is the tool that caused dramatic change in individual and collective behavior during the last 20 years and which is a manifestation of many of the ideas of the pervasive computing paradigm: the mobile phone. In this work, the mobile is regarded as a proxy for people. Through it, the social environment becomes accessible to digital measurement and processing. To understand the large amount of data that now becomes available to automatic measurement, we will turn to the discipline of social network analysis. It provides powerful methods, that are able to condense data and extract relevant meaning. Visualization helps to understand and interpret the results.

This thesis contains a number of experiments, that demonstrate how the automatic measurement of social proximity data through Bluetooth can be used to measure variables of personal behavior, group behavior and the behavior of groups in relation to places. The principal contributions of this thesis to the named research field are:

- A methodology to visualize personal social context by using an ego proximity network. Specific episodes can be localized and compared.
- A method to compare different days in terms of social context, e.g. to support automatic diary applications.
- A method to compose social geographic maps. Locations of similar social context are detected and combined.
- Functions to measure short-term changes in social activity, based on the distinction between strange and familiar devices.
- The characterization of Bluetooth inquiries for social proximity sensing.
- A dataset of Bluetooth sightings from an ego perspective in seven different settings. Additionally, some settings feature multiple stationary scanners and Cell-ID measurements.
- Soft- and hardware to capture, collect, store and analyze Bluetooth proximity data.

1.4. Structure of this Thesis

The content of this thesis is organized into four parts and twelve chapters. In the first part, we set the frame of this work which we call the urban pervasive infrastructure and examine related work. The next part introduces the software and hardware used for the measurements of proximity data and verifies its parameters in the lab and in the city. The third part describes the collection of the dataset and presents the methods we developed to analyze and visualize social context. The thesis concludes with the fourth part.

Chapter 1 gives an introduction and briefly outlines the contributions and the structure of this work.

Chapter 2 takes a view on urban computing. The development from digital to sensor cities is described. The concept of the *urban pervasive infrastructure* is introduced as a framework to understand and design pervasive computing applications for the places between home and work. Characteristics and metrics of this infrastructure are outlined. The chapter is based on the article "Understanding and Measuring the Urban Pervasive Infrastructure" by Vassilis Kostakos, Tom Nicolai, Eiko Yoneki, Eamon O'Neill, Holger Kenn and Jon Crowcroft [93].

Chapter 3 undertakes an examination of the concept of social context. We understand the concept from the perspective of various sciences (incl. sociology and psychology) and give a special focus on its relation to urban environments as well as the meaning of proximity for human relations. Pervasive computing research is reviewed for its usage of social context and its applications are classified. We close the discussion with an examination of tools to analyze and visualize social context, including social network analysis.

Chapter 4 discusses the technical options to measure the proximity of people in the city. State of the art technologies are systematically reviewed. Bluetooth device inquiries prove to be the best choice for the measurement of social context through physical proximity. Relevant technical details of the Bluetooth discovery specification are reviewed.

Chapter 5 describes the WirelessRope—a system of hard- and software components we developed to measure physical proximity by Bluetooth. It consists of three tiers, mobile, stationary and server, which provide a flexible structure for deployment in various settings. Its most important part for our study consists of a couple of mobile phone programs, which were used to collect the largest part of our dataset.

Chapter 6 evaluates the performance of the WirelessRope and Bluetooth for the purpose of our study. First, we verify the general parameters of device inquiry in the lab. Then, the impact of different environments (corridor or open field) on inquiry performance is evaluated. We also examine a number of distance indicators.

Chapter 7 takes the WirelessRope out into the city, to measure, how many percent of people can be detected by the Bluetooth measurements. We sample three locations in Bremen, Germany, and one in San Francisco, US, and compare them to a former study in Bath, UK. The chapter is based on the paper "About the Relationship Between People and Discoverable Bluetooth Devices in Urban Environments" by Tom Nicolai and Holger Kenn [135].

Chapter 8 outlines the collection of our dataset, which contains a sample of the social context of our proband—in the form of Bluetooth proximity data. It was collected over the course of several months in seven different settings in different parts of the world. Separate settings are augmented by additional data, containing Bluetooth scans from a set of stationary scanners or Cell-ID data. We begin the analysis of the dataset by composing an ego proximity network and measure its basic properties. This network is a way to map social context as a whole. Episodes can be localized on this map and set into relation.

Chapter 9 focuses on temporal elements, namely days and sets of several days, within the dataset. We develop a method to compare such temporal entities on the basis of social context. Applying the method to our dataset, we find connections between all the days of our set. Since the subsets were collected on entirely unconnected events in space and time, this indicates the inherent connectedness of our concept and that the few percent of people we can detect by Bluetooth are enough to connect the various settings.

Chapter 10 adds another component to the analysis: location data. By combining social proximity with geographic location, we are able to localize social networks on a floor plan and street map. We present a method to condense the information on this map and create blocks of geographic areas, defined by their social context as well as groups of people, defined by their appreciation of specific places. This chapter enables the construction of social maps, which may show us not only the shortest way, but a way we may enjoy for the people we encounter on the route.

Chapter 11 examines a six-day dataset on a finer scale. We present two functions, which indicate changes in the social environment and calculate them for every five minutes of the data. The functions are based on the distinction between strange and familiar Bluetooth devices. This method was first presented in the papers "Exploring Social Context with the Wireless Rope" by Tom Nicolai, Eiko

Yoneki, Nils Behrens and Holger Kenn [136] and "Towards Detecting Social Situations with Bluetooth" by Tom Nicolai and Holger Kenn [134].

Chapter 12 provides a detailed summary of the principal contributions of this work and outlines an agenda for future research into social context based on the experiments and interpretations described in the previous chapters.

2. Urban Computing and the Structure of Cities

While much computing research has been concerned with home or work, the focus has recently been shifting toward "third places"—the spaces between home and work ([172], p. 36).

Urban computing—or public computing, as Shklovski and Chang call it [172]—deals with the use of pervasive computing in public, urban spaces. People, architecture and technology together provide the *urban pervasive infrastructure (UPI)* that urban applications use. Understanding this infrastructure is crucial to developing such applications. In this chapter, background information about computing in urban environments with a focus on social interaction is given, metrics for understanding the urban pervasive infrastructure are described, and a set of observation, analysis and simulation methods for capturing and deriving those metrics are elaborated on.

2.1. Digital Cities

The idea of urban computing is heavily influenced by concepts from pervasive computing being grounded on mobile devices. It also developed out of the so-called *Digital Cities*. These Web-based spaces are especially developed for the citizens of a specific city, for tourists planning their visits, and others related to the place. Digital cities usually provide their users with information about local cultural and social activities. They often comprise functions for local interaction, for example with digital black boards that can be used for trading or help to find an apartment. In the USA, AOL is running a series of such sites [4], in other cases they are supported by the local government, like Bremen, Germany [19]. The Digital City Amsterdam (DDS) evolved from an initiative to democratize access to the Internet. In the beginning, it was build on BBS technology, later a Web-based interface was added. A map with different districts, that did not match the real layout of Amsterdam, was implemented to create feelings of neighborhood among the citizens. However, as a study about the Digital City Amsterdam (DDS) shows [190], the users of a digital city are not necessarily as heterogeneous as the citizens of a city. Moreover, the users of DDS are distributed over the whole of the

Netherlands, and not constrained to the capital. Thus, the connection between real and virtual place dissolved over time, giving the DDS its own, distinct identity.

In his work about the Digital City Kyoto, Toru Ishida developed several ideas to form a strong connection between a real city and its virtual counterpart [79]. His idea was to create a complement to the real city, that could only exist in the connection with it. In contrast to the DDS, Digital Kyoto was based on a map of the real city. Websites related to the city were connected to the map automatically by a program searching the Web for pages with corresponding address information. In addition to a 2D map view, a 3D view was also part of the system, to provide a detailed view from a visitor's perspective. A very strong tie between virtual and real city was created by the deployment of sensors in Kyoto that were connected to the website. About 300 traffic sensors provided data about the city buses. The differences between the buses' schedules and their real positions could indicate traffic jams. There were other sensors to capture weather conditions as well as cameras to stream live video. While especially the traffic information was irrelevant outside of the city, it was so interesting when navigating through the city that it inspired mobile interfaces for access on the move. There was also a strong focus on social interaction in Digital Kyoto [80]. Besides traditional community building tools like newsgroups and chat rooms, it was imagined to localize citizens in the city, e.g. by GPS, to provide better, localized services.

There is no clear definition of a digital city. Gumpert and Drucker emphasize the aspect that public information is being transmitted electronically in a digital city. Thus, there might be terminals providing one-stop public access to local information [60]. A *teleport* might also be part of a digital city. This is a combination of real and virtual space, in the sense that office and living facilities are equipped with broadband telecommunication systems. Often, teleports are a result of direct demand by the local industry (e.g. in New York), in other cases they are initiatives to improve the attractivity of cities for certain industries (e.g. in Bremen) [59]. However, teleports do not focus on specific applications. Rather they provide the pure infrastructure and are neutral to the applications of their users.

Not so much focused on industrial use are several urban community systems in developing countries, e.g. in South Africa or Mexico. In these cases, the goal is to embed communication technologies into the social life of the population [53]. These systems are designed to support social interaction and strengthen local communities. They are often created together by the local government and the industry. The Urban Tapestries project is another example to support a participatory approach in urban life, fostered by wireless communication and mobile computing devices [105].

2.2. Sensor City

Kryssanov et al. take a semiotic view of the digital city to understand the underlying principles [98]. Departing from the metaphor of the city, they distinguish the concept of the digital city from static websites containing information about a city. They follow the argument of Portugali, who writes that cities are complex, self-organizing systems, and not rather static configurations [153]. As such, sudden and unexpected changes in the structure of a city might happen, e.g. as a result of changes in the market or in employment opportunities. These changes happen on a major level, as well as on minor levels. Thus, a more or less static website about a city existing in isolation from the city might not be suited to reflect all the subtle processes happening inside. If such websites only provide a thin and controlled interface to feed data into, they are running the risk of being outdated and thus meaningless to citizens in most situations. The separation between website and city might also be physically evident. It might run on servers anywhere in the world without any embodiments present in the real city.

Portugali further describes a city of being perceived both physically and cognitively by the citizens. In this model, the locations and actions of individuals in the city are determined by their cognitive maps and in turn affect individuals' cognitive maps of the city. The digital part of a city might directly integrate in this model. It also affects the cognitive maps of the citizens, which, again, has an effect on the city itself through the actions of its citizens. In practice, the effect the digital counterpart has on the physical structure of a city has not been studied, yet. The complexity of a city and the subtlety of the influence of the digital make such an attempt very difficult.

Going back to the work of Kryssanov et al., they reinforce that "the users together with their knowledge can and in fact should be considered as indispensable and *constitutive* parts of the digital city" ([98], p. 60). Thus, the users interpret the input they receive from the digital and make sense of it to navigate the physical city. To close the loop between the physical city, the digital city and its users, a broad interface is necessary to facilitate the information flow from the physical to the digital city. As argued before, a thin and controlled interface, such as an administrator's part of a content management system, runs the risk of not being responsive enough, thus dissolving the unity of digital and physical city. Recently, there have been several experiments to use sensors spread throughout the city to close the loop with a primarily automatic system (see figure 2.1).

This loop between digital city, physical city and its users is evident in several existing systems. E.g., the prototypes described in [82] exhibit this behavior on an artistic level. The first prototype called "Audio Tags" contains a proximity sensor, a microphone and a speaker to interact with the physical city as well as with users. When a person approaches the device, it replays a previously recorded message and allows

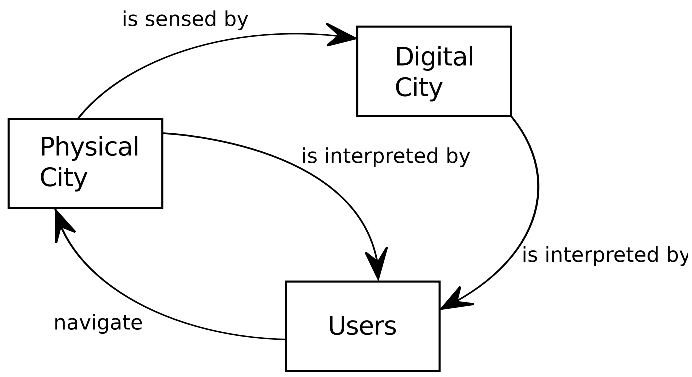


Figure 2.1.: Closing the gap between physical and digital city with sensors

the user to record a new one being replayed to the following users. It thus senses its environment and gives information to be interpreted by the users. As a result, the users might change their behavior in some ways. The second prototype “Glitch” works on the same principle, but contains different sensors to act on electromagnetic interferences of mobile phones.

While these prototypes are isolated and not connected to a larger system, the London congestion pricing system consists of distributed sensors and interfaces, all connected through a digital network [107]. Since 2003, it is installed in the center of London to reduce traffic congestion in this area and raise revenues to fund improvements in transportation. Drivers are required to pay a fee before they enter the center of the city with their vehicles. There are several interfaces to do so, including a website and a mobile phone service. Vehicles in the controlled area are identified by video cameras and the system automatically checks, if the fee was payed. If not, a fine is assigned directly. As anticipated, the system changed the behavior of the citizens towards using other kinds of transportation than private vehicles and thus reduced traffic congestion.

Overall, there are a lot of different kinds of sensors in place in contemporary, modern cities, including video cameras, movement sensors, pressure sensors and RFID systems. Even mobile phone signals are used to drive applications beyond their original function of enabling phone calls (see table 2.1). Bluetooth sensing has received attention lately, especially as a means of augmenting poster advertisements with digital content, like mobile phone ring tones. The principles and applications of Bluetooth sensing are presented later in this work in detail, since Bluetooth is used extensively throughout the conducted experiments.

Another interesting aspect of urban sensor usage is the sensors’ mobility. While most sensors today are statically attached to or integrated into the built urban architecture, mobile sensors, affixed to vehicles or citizens, provide advantages under certain circumstances. Liu et al. found out, that a greater area can be covered by less mobile

Sensor	Application
Video camera	Security, traffic control
Pressure sensors	Trigger traffic lights
Movement sensor	Light control
RFID	Theft control, shopping
Mobile phone signals	Mobile marketing, traffic control
Bluetooth	Mobile marketing (posters), tourist information

Table 2.1.: Examples of sensors embedded in contemporary cities and their applications

sensors than by stationary sensors [108]. The downside to this approach is, of course, that locations are only covered part of the time and not constantly. Furthermore, data transmission in networks of moving sensors is difficult, because of the limited and opportunistic contact times between sensor nodes. Hui et. al researched this problem with sensors carried by visitors of a conference and characterized the contact times of the sensors [77].

Within the MetroSense project [122], Campbell et al. propose a city-wide system of combined static and mobile sensor nodes to provide both, good connectivity and good coverage, for sensing applications [24]. Therefore, they developed a three-tier architecture consisting of a mobile sensor tier, a gateway tier, and a server tier. The mobile sensors in their system are carried by people or affixed to vehicles. Besides the capability of sensing the environment, they can also sense each other, thus detecting contacts between people, between people and objects and just between objects. The devices of the gateway tier are stationary and combine the functions of the mobile sensors with the additional capability to provide constant access to the server tier. The server tier contains various servers that run various applications and can be assumed to have large storage and computing power. Applications on the mobile sensors are expected to work without the ability to control their mobility. Thus, the hosts move freely and the sensor nodes take advantage of that movement. Among others, the authors imagine an application—called "Pulse of the City"—, which takes the locations of people together with the density of people at that place and additional manual input, to provide an up-to-date view of peoples' activities in the city.

The Pervasive Mobile Environmental Sensor Grids (MESSAGE) project [124] aims to collect data at a metropolitan scale through mobile phones carried by cyclists, cars and pedestrians monitoring carbon dioxide values to control traffic in the city of Cambridge, UK. Similarly, the urban sensing project at CENS [189] seeks to develop cultural and technological approaches for using embedded and mobile sensing to invigorate public space and enhance civic life. Another approach to a network of urban sensing devices is presented by Riva et al. [160]. There are also consumer oriented sensing applications such as Nike+ [139], sensor-enabled mobile phone applications, health related sensing, and sensor enhanced urban gaming [32]. The potential of the research field of large scale sensor networks is also reinforced by initiatives like Sensorplanet [170], which is backed by Nokia and is aiming to provide test platforms as well as a data repository for sensing applications.

Although in many of the applications and architectures presented, the sensors may operate automatically without assistance of the person carrying the device, there are several approaches to include the users and to empower them by the sensors. Massimi et al. experiment on this theme with a scavenger hunt game, in which the players must cooperate through sensing devices to solve a quest [118]. The "Freeporter" system enables the users to create reports about events with mobile phones while they are happening outside, to distribute them over the Internet, and also to receive news updates instantly, filtered through a personal reputation system [133]. Mann even conducted work on embedding tiny video cameras into ordinary eye glasses to enable applications like personal, life-long video documentary [115]. Rheingold gives many examples of spontaneous self-organization through mobile devices in his book Smart Mobs [158].

The mentioned development efforts targeting a city-wide scale might be tracked back to smaller experiments conducted in smart room environments. E.g., Elrod et al. installed activity sensors in office rooms to save power by automatically controlling lights and heating [49]. The PARCTAB mobile devices were enabled to act as manual controllers for the rooms. Pentland managed to have computers recognize humans' gestures and facial expressions through video cameras and used this information to control computers embedded in the environment [150]. While these first efforts were focusing on the interaction of one user with the environment, McCarthy's as well as Sawhney et al.'s work extended the scenario to halls and gangways used by several people [119, 162]. Research on urban computing has again moved on to scenarios being more complex as they are used by even more people with different intentions. Thus, urban computing is concerned with third places, those spaces between home and work.

2.3. The Importance of the Urban Pervasive Infrastructure

No two cities, or different places within the same city, are identical. Cities within a country can be as diverse as cities in different countries. The range of complex factors making a city unique includes that city's urban spatial form, the people who inhabit it together with their culture and values and the technologies that operate in it. These factors may be considered as part of the infrastructure of the city. In designing urban computing systems, it is essential to take account of this infrastructure.

Mainwaring et al. conducted a study in three global cities—London, Tokyo and Los Angeles—that compared the "mobile kits," the things people are carrying around with them, of the cities' inhabitants. They were looking for a common basis for the design of one urban interface. What they found were many commonalities in the usage of mobile devices, even though there were clear cultural differences [113]. In contrast, Williams and Dourish emphasize the differences between cities. They object that a city is a generic setting and that findings in one city are automatically applicable to any other city. Especially, the cultural and historical background of the city, as well as the background of the "user" of a city, make a difference [201].

So, how can those differences, or similarities, be expressed in ways that are meaningful and useful to the designers of urban applications? Just as traditional desktop-bound applications utilize technological infrastructure for their operation (e.g. networks, software services, etc), urban applications can draw on the available urban pervasive infrastructure.

For designers, an understanding of the urban pervasive infrastructure can be useful to resolve design questions about the types of applications that may be built on this infrastructure. For example, which cities, or parts of a city, would be best suited for deploying a specific urban application? Can applications be optimized, based on the understanding of the infrastructure and its affordances?

The next sections introduce metrics and methods constituting a basis for specific research. However, before delving into describing the methods for observing and analyzing data, the concept of the urban pervasive infrastructure and its importance is described. Then, a set of concrete metrics that are used to measure and understand this infrastructure is given.

Previous work has shown particular components—human (e.g. [46]), technical (e.g. [28]) and spatial (e.g. [71])—of the urban pervasive infrastructure to be important. It may be beneficial to draw on the lessons of this disparate work. Furthermore, a richer understanding, and more successful system design practice, may be realized by taking a holistic approach that integrates related disciplines and projects in a systemic view of the urban pervasive infrastructure. Viewing the city as a system, the elements of people, space and technology combine to constitute an urban pervasive infrastructure over which urban applications may be deployed.

A key requirement for studying the UPI is capturing trace data of the real world (e.g. human mobility, intermittency of connections between people) in order to construct realistic synthetic models. For example, the Reality Mining project [157] collected proximity, location and activity information, with proximity nodes being discovered through periodic Bluetooth scans and location information by cell tower IDs. Several other groups have performed such studies [46, 7, 25, 121]. Most of these, such as [7], used Bluetooth to measure mobility, while others, such as [25] and [121], rely on WiFi. The duration of experiments varies from two days to over 100 days and the numbers of participants vary from eight to over 5,000 (see the Haggle project [61]). The Crawdad database [34] provides extensive traces, which are useful for the validation of forwarding algorithms and routing protocols that operate through learning characteristics of node mobility.

However, previous research lacks an integrated approach that considers the various aspects of the UPI—people, spaces and technologies—as a system. Examining aspects of the UPI in isolation, even when large datasets are available, can provide results that are not easily transferable to new settings. On the other hand, considering the UPI as a system gives a more integrated picture of a city and provides the foundations for an integrated approach to build urban applications and services. This allows for the correlation of findings from various cities, and the transferring of those findings.

A number of instances can be considered, where an understanding and modeling of the UPI can produce better or new applications. For example, previous research on GSM positioning for mobile phones highlights the need for detailed maps of cell tower IDs and reception in urban areas [28], which are essential elements of the UPI. Apart from determining an exact geographic position, it might also be interesting for a user, that a system remembers places that are meaningful for a user—such as “home” or “Tony’s Pizzeria” [67].

In addition to location, the UPI can provide information about a user’s social context. Social network analysts typically use questionnaires and interviews to investigate social networks. Shortcomings of this method are that it is resource and time consuming, longitudinal data collection is difficult, and the data is biased by self-report errors. The Reality Mining study at MIT involving one hundred users of mobile phones running a Bluetooth scanning application has shown that it is possible to automatically derive affiliation networks and to model friendship relationships from the scan data [46]. Moreover, this data is not subject to the shortcomings noted for the traditional questionnaire and interview methods, despite the numerous technical issues that may arise during a study. So far, such studies have been carried out in a controlled environment considering only contacts between study participants. When merged with an understanding of the UPI, such studies can be extended beyond this controlled setting and related to the wider social context. More crucially, however, an understanding of the UPI can help localize applications, which can then target different cities more effectively.

In addition to localization, understanding the UPI can help with the evaluation. Urban applications are often difficult to evaluate. This is especially true if they are designed for opportunistic events or require a certain critical mass of new devices. From a usability perspective, it is common to conduct tests in a manageable setting, or to capture users' opinions in focus groups, interviews and questionnaires. Models based on real-world measurements of the UPI can be valuable additional evaluation tools, saving considerable resources and providing helpful directions at the start of a project. By analyzing the UPI, it becomes possible to identify a priori settings and communities where potential applications are better suited. Many systems can benefit from this type of analysis, such as the "Augmented Word-of-Mouth" [75], that leverages an epidemic approach to forward messages to people based on physical proximity.

Finally, modifications and extensions to the UPI can benefit from an understanding of its structure and internal workings. For example, architects and city planners use tools like space syntax [71] to model existing cities and design new ones. In addition to physical architecture, the habits of the inhabitants, such as the routes they take, are also important. With this knowledge, urban applications can be optimized for the characteristics of a specific urban context. Additionally, urban simulators can take advantage of this data to make more accurate predictions of pedestrian movement and flows (e.g. [187]). This in turn will allow for better lab-based evaluations of proposed changes to the structure of the city, examples of which are the addition of bridges, demolition of building blocks and redesign of urban landscapes. Furthermore, plans for the redevelopment and deployment of technologies can be assessed in a similar manner, taking into account the concepts and metrics for each city's UPI. For example, the installation of wireless access points can be informed by the spatial structure, the patterns of pedestrian movements which result in expected bandwidth requirements, and even knowledge of the types of mobile devices in the city.

The premise is, that a systemic understanding of the UPI can help to develop urban applications that play to the strengths of this infrastructure. Such an understanding, however, requires the establishment of clear concepts, metrics and methods.

2.4. Characteristics and Metrics of the Urban Pervasive Infrastructure

Before describing methods to deal with the UPI, a set of characteristics is identified. In this section, these characteristics are introduced, along with metrics and explanations of their use. Of course, there are potentially infinite aspects of a city to be studied, however here the focus is on those aspects that available technology permits and for which adequate datasets can be captured and analyzed. The following characteristics of the UPI are investigated:

- mobility,
- social structure,
- spatial structure,
- temporal rhythms and
- Facts and figures.

Mobility is a key feature of both humans and technology [7]. Each city has a unique pattern of mobility. Considered from an egocentric perspective, useful metrics are distance traveled and speed. When considering mobility from an exocentric perspective, flow becomes a useful metric (people/hour), as well as visit duration (in the form of a time-based distribution). Mobility itself can also be seen as the amount of randomness or entropy in a city.

Social structure describes social groups, social behavior and patterns of encounter. Social structures can be examined from an egocentric or exocentric perspective and involve issues like group size, number of singles vs. couples, etc. Concrete metrics can be adopted from traditional social network analysis such as degree, betweenness and closeness [95]. Measuring the social structure in a city is vital to understand patterns of behavior that are observable on an aggregate level.

Spatial structure gives insight into aggregate behaviors and patterns observed in a city. Space syntax provides tools to examine the city from a purely structural perspective and to compare cities and sites within a city in terms of structure. Concrete metrics for spatial structure include integration, choice and intelligibility. Spatial structure has been shown to affect various high-level human behaviors such as shopping patterns and crime [70].

Cities and people have their own *temporal rhythms*: daily, weekly and seasonal. An understanding of these, along with concrete measures for comparisons is important to understand the urban pervasive infrastructure. Typically, cities' temporal patterns are affected by laws and restrictions (e.g. pubs must close at 11pm), work schedules (at the daily and weekly scale) as well as seasonal variations such as the weather and holiday seasons. Concrete metrics of such rhythms can be expressed as time-based distributions (see [7] and [145]).

Finally, *facts and figures* refer to any statistical characteristic that is applicable to people, technologies and spaces. For example, facts and figures about humans can be how many people go clubbing, or how many teenagers live in a city. A technological characteristic can refer to the spread of WiFi or Bluetooth. An architectural characteristic is the number of parks or restaurants in a city. Facts and figures are obtained by applying established empirical methods such as surveys, by consulting maps and census data or are recorded by technological means.

2.5. Methods for Understanding the UPI

In this section, methods are described that are applied to gain insight into the concepts described above. There is no one-to-one mapping between the methods described below and the concepts of the UPI, and in many cases, combinations of methods, through observation, analysis or simulation, are used to generate results. For example, to understand mobility, various observation methods may be used to gather data in combination with one or more of the analysis methods described in this section.

2.5.1. Observation Methods

A challenge faced is recording, representing and understanding the patterns of mobility and presence in our cities through the use of pervasive technologies. Most wireless technologies have characteristics that render them appropriate for study by this methods. For instance, the vast majority of Bluetooth devices, such as mobile phones, have a relatively short range and map very closely to the movements of people around the city. In contrast, typically static WiFi or GSM access points can be used to identify locations in a city, while the signals emitted by WiFi devices can be related to both static and mobile devices such as desktop and laptop computers.

A common observation method used to capture aspects of the UPI is *wardriving* [78]. It involves systematically moving about a city to record various detectable or visible features of technology. This includes WiFi and Bluetooth activity, the presence of mobile phone masts, the use of mobile phones and cameras, all of which produce maps (see [200] for sample WiFi maps) with color-coded information about the presence or levels of activity of certain technologies. Additionally, physical aspects of the city itself can be recorded in maps highlighting features such as parks, schools, graffiti, and housing vs. commercial areas.

A further observation method is the *augmented gatecount* [145]. Gatecounts are used to establish the flows of people at sampled locations within the city. A gate is a conceptual line across a street, and gatecounts record the number of people crossing that line. The observer counts the number of people crossing the gate in either direction. This process can be augmented by providing the human observer with equipment that monitors the presence of technologies, e.g. by Bluetooth inquiries. Additionally, the observer may manually record technology related behavior such as the number of people using mobile devices like phones or cameras. This method provides data correlating the presence of a technology (e.g. Bluetooth) or behavior (e.g. use of mobile phones) with the local population.

To observe the open spaces of a city (outside, such as a plaza, or inside, such as a café) *augmented static snapshot* [145] may be used. A human observer manually records human activity, including apparent technology use, while simultaneously recording technology use with appropriate scanning devices. The method is used to record both stationary and moving activities, and is particularly useful when directly comparing the two types of space use. This method highlights the different types of space use in an urban area. It gives an understanding of how people visit and use a particular space, and how these habits bring people into contact with each other. For example, it may be observed that a seating area in a park is actually not used for seating but for playing by children. A common observation is the use of certain spaces by people making calls on their mobile phones or using their laptop computers, and the way these people locate themselves with respect to their surroundings and other people.

People's mobile devices, when used as mobile scanners, can capture a personal view of the UPI. Focusing on the personal perspective provides an understanding of the contexts and habits of individuals. To achieve this, participants must be instructed to interact naturally with their environment during the measurement. Depending on the aspect of interest, different scanning technologies can be utilized. For example, GPS gives insight into spatial behavior while Bluetooth scanners emphasize social behavior.

The above methods offer longitudinal data, too, by installing the scanning equipment for long periods of time (e.g. [145]). In this case, there may be no human observations to correlate with the data, however such long-term scans can provide richness in terms of patterns of the city over time and relationships between people. This is especially true when combining data from multiple locations, as well as combining data from mobile scanners and stationary scanners. As part of the Cityware project [31], a Bluetooth based infrastructure was designed and implemented consisting of various components to combine these observation methods in a single system. It was installed on a long term in the city of Bath, UK.

2.5.2. Analysis Methods

In the previous section a number of observation methods were described. Here, it is shown how to analyze the data from these observations. Analysis of wardriving data is quite commonplace. It is used to indicate areas of interest as well as patterns of behavior and use over time. Similarly, facts and figures can be calculated using statistics tools, depending on the exact facet of the UPI in question. For instance, a city's WiFi coverage can be calculated by analyzing wardriving data.

The majority of the analyses described in this chapter are focused on gatecount and static snapshot datasets gathered in the city of Bath, UK. Analysis of the gatecount datasets allow to identify interesting mobility and temporal patterns, as well as facts and figures about the UPI. First, gatecount datasets are used to infer patterns and trends in the movement of people across the city. Patterns are observable on many scales, from hourly to seasonal. Additionally, it is possible to identify facts and figures, such as the overall penetration of Bluetooth in a city. Specifically, in Bath it was found that about 7.5% of pedestrians carry mobile phones with Bluetooth set to discoverable mode [145]. Furthermore, such data can be used to identify device classes, or indeed device brands. Knowledge of the actual mobile devices in a city (such as their brand and operating system) may be an influential factor for the development of applications.

A further focus of work has been the analysis of long-term data captured in static snapshot locations. Based on the co-presence of discoverable Bluetooth devices in a location, people's encounters in space can be inferred [95]. The data can be represented as social network graphs (see figure 2.2), linking persons that encountered each other. In this example of data from a pub in Bath, the size of nodes represents the amount of time those devices have spent in the pub, while their color represents their betweenness[1] (red: 1, blue: 0). The length of edges is determined by the graph layout algorithm and does not relate to any specific properties. These graphs are then suitable for traditional complex network analysis. The presence of power law distributions is present in these graphs [95], which are indicative of self-similar, real-world networks. Such distributions, which can be found in earthquake magnitudes, word frequencies, city sizes, and the structure of the Web, open up several possibilities to apply established analysis techniques to the datasets. Furthermore, by adjusting the rules used to derive the graphs, it is possible to focus on different aspects of a city. For example, devices that appear and disappear together may be emphasized, indicating possible groups of people and thus social ties. This allows further to infer communities within the city.

The combination of multiple static snapshots or gatecount datasets provides useful insights into trails and patterns of movement. For instance, [147] have analyzed a WiFi dataset for trails, or hops, between various locations in the city. These show people's movement through the city in terms of their connection to WiFi hotspots. This type of analysis provides insights into questions like "Which trail in the city is mostly followed on Friday evenings?", which in turn can shape the design of urban applications.

[1]Betweenness is a notion of social network analysis. It quantifies the importance of nodes for the diffusion of information throughout a network. For a definition see [193].

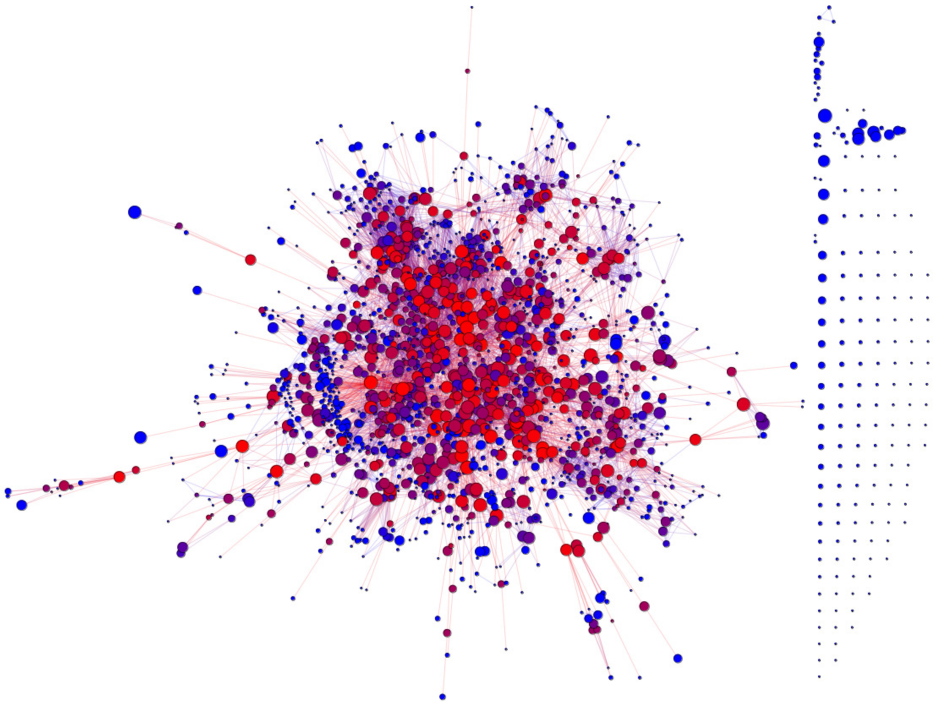


Figure 2.2.: Social network describing encounters of devices within a pub in the city of Bath [93]

Another useful technique is space syntax [71]. It models the structure of cities and its effect on pedestrian movement. This analysis is done in two steps. First, ordinance survey maps are used to analyze the spatial structure of a city, purely in terms of lines of sight in the open spaces such as streets. This results in predictions about which streets are likely to be busy and which are likely to be quiet. In the second step, observation data of the actual pedestrian flows are compared to the predictions. In this step, predictions are fine-tuned by changing the weighting on different variables used in the predictions. Thus, using observation data as a guide, space syntax identifies the important variables that can be used to accurately model pedestrian flow. Knowledge of these variables allows for more accurate explanations of the spatial dynamics, as well as more accurate predictions of the effect of space on behavior.

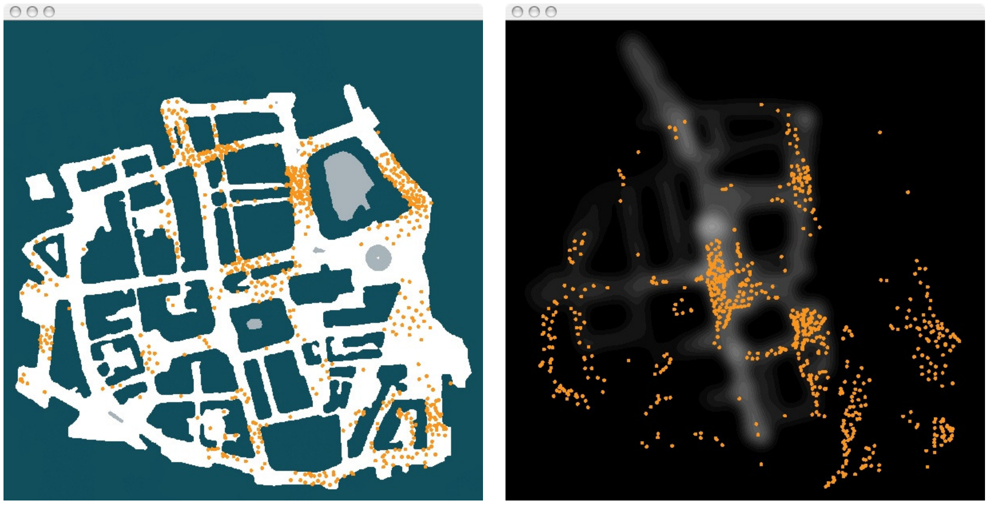
Finally, device contact patterns, such as contact duration and inter-contact duration[2] are used to study ad-hoc network opportunities that arise in a city. The analysis of data from static snapshots recording Bluetooth traffic has uncovered inter-connection patterns and has been used to develop data forwarding algorithms [25]. Specifically, the distribution of inter-contact time follows an approximate power law over a large period of time. Inter-contact durations are of particular importance because their distributions determine the viability of forwarding algorithms, as shown in [25]. Additionally, temporal graphs can be used to determine admissible and optimal paths through the multitude of devices in a city's UPI. Furthermore, forwarding algorithms can consider the levels of clustering in pedestrians' movement and the affiliation networks in a city.

2.5.3. Emulation and Simulation

A benefit of augmented gatecounts and static snapshots is that they produce time-stamped records of events that can be used for replay in sequence. By emulation, "what-if" situations can be examined, and the effects of different technologies or different circumstances can be studied. In emulation, the diffusion patterns of information through the social networks derived from the analysis of static snapshot, can be studied by testing different types of rules. For example, it can be considered how a small (1KB) and a large (1MB) application spreads through the city, based on recorded device encounters. Further, inter-connection times can be replayed in order to adjust the forwarding algorithms. Emulation can act as an initial testbed for many applications, where facets of the pervasive infrastructure can be brought into action inside the lab.

Having a lab testbed is important, as working and observing in the city is expensive, both in terms of money and time. For instance, installing and maintaining long-term scanners requires equipment, bandwidth, and personnel time. Furthermore, it is not always possible to install scanners in desired locations. For these reasons, observational

[2] The duration between two successive direct contacts between the same pair of nodes.



(a) Street map of the center of the city of Bath (b) “Bluetooth map” of the same area

Figure 2.3.: Snapshots of a city simulation [93]

datasets may be extended by simulations. Simulations can generate large amounts of data inexpensively, but usually with less confidence in their validity than that of empirical data.

The most common mobility models used in simulation for mobile ad-hoc networks are the *random walk mobility model* and the *random waypoint mobility model* [22]. Both simulate node movement in a rectangular area. In the *city section mobility model* [22], nodes move on streets choosing destinations at random and follow the shortest paths to them. However, these mobility models rarely reflect accurate real world situations. Yet the use of real world traces is important, albeit often difficult to obtain. By taking the cognition of agents into account, it is possible to generate a more realistic behavior, as shown by Turner and Penn [187]. Comparing simulated agents’ behavior with real, observed data, provides a measure to a simulation’s quality.

A robust simulation of a city is very useful to inform urban planning [37]. On the one hand, planners can test their plans to see if any unwanted and unforeseen behaviors or side effects might arise. On the other hand, such simulations let citizens explore consequences, too, and let them develop and test alternatives. As such, the process of urban planning might develop more democratic characteristics.

An interesting question is how the existing simulation models provided by space syntax can be optimized. These models simulate pedestrian movement in the city, and effectively would allow one to flood a (simulated) city with mobile agents and information packets. Figure 2.3 illustrates such a scenario. The white areas on the right map indicate high Bluetooth activity. The yellow dots indicate information packets that move about the city via Bluetooth.

An important step is to change the properties of the agents' cognition [187] to match our observations of flow, encounter, and interconnection times through the (real) city. Once a good fit between the observational data and the simulation data is achieved, simulations can be used as an additional source of data. For example, virtual gate-counts and static snapshots within the simulation could be carried out, resulting in a large dataset to augment field observations. At the same time, it is important to use computer simulation to validate the model derived from the real world.

2.6. Privacy Considerations for the UPI

In the introduction (see section 1.1), we discussed a defining property of cities compared to small rural villages: anonymity. Urban computing is situated in public spaces, such as streets and squares. Urban anonymity is based on the vast amount of people at a place, where no person can keep track about each one of them. However, urban computing has the potential to challenge this anonymity in a variety of ways. We would like to point out some of these threats.

Privacy is already challenged by several companies and systems related to the services they provide. Elcoate et al. mention

- mobile telephone network providers,
- banks,
- credit card providers,
- Internet Service Providers and
- workplaces (often through swipe card entry to buildings)

that track the behavior of their users [48].

Advertisements, delivered on mobile phones and triggered by the location of individuals, even threaten to transform private walks through a city. Curry et al. describe such systems as being a very distractive technology [35]. But moreover, technology may break the city's anonymity by facilitating what no single person might be able to do: watching a person's every step through the city. Surveillance, especially by closed circuit television, is already high in several cities. E.g., in 1999, an individual was already captured on 300 different cameras from 30 different organizations in London on a typical day [60]. Together with face recognition software and algorithms that recognize the specific movement patterns of individuals even when their faces are not visible or covered, anonymity gets compromised.

A similar threat emanates from Bluetooth sensing, too. Of course, people do not respond to Bluetooth inquiries, but their devices (e.g. mobile phones, PDAs and computers) do, when set to "visible mode." Especially the location of peoples' mobile phones is a good indicator for their location in public spaces, since they typically keep their phones close to them. Bluetooth, like any other networking technology, incorporates addresses that are unique to each device as a basis for communication. Generally, these addresses are fixed and cannot be changed. Thus a person's presence can be detected for as long as he keeps his mobile phone. Usually, phones are only changed every two years, which is the typical extent of a contract with a mobile operator. Although there is no direct connection of a phone's Bluetooth address with its owner's identity, this connection could easily be established by a phone company or by a combination with other technologies like video surveillance. The only way to ensure privacy seems to be by disabling "visible mode," thereby limiting the usages of this technology. Others have noted this limitation, too, like Huang and Rudolph [74], who demonstrate a Bluetooth-based location system, that preserves the privacy of its users. However, their approach comes with increased battery usage and memory requirements on the mobile devices, compared to an approach that does the computation in the environment, and not on the devices of the end users [3].

Kostakos et al. developed a conceptual framework to analyze privacy issues in pervasive computing [96]. They propose a classification of places and uses of pervasive technology as being

- private,
- social or
- public.

They agree with Palen and Dourish, that privacy is under continuous negotiation, depending on the environment and the behavior of the surrounding people [146].

In terms of their framework, Bluetooth provides the options of being "private" or "public," depending on Bluetooth's "visibility" setting. Thus, users have the option of denying contact by others completely or to be open for contact by anybody in range. This situation has spurred misuses like "Bluejacking," where text, picture, or sound messages are sent to unknown people. Victims may get confused or might think their phones got hacked, although it is generally not harmful. A function relating to the "social" option, where only trusted persons or their devices were allowed to detect a device and to connect to it, is missing in the Bluetooth specification. Even worse, if the address of a device is known to an attacker, he can detect it, even if the victim's

Bluetooth device is set to "invisible mode."[3] In relation to the Bluetooth characteristics of the UPI introduced in this chapter, the "social" option would empower people to consciously select the services in which they would like to participate. With current Bluetooth technology, such a distinction is not possible.

The MetroSense architecture described earlier, basically suffers the same problem of not providing location privacy. It includes several solutions for other privacy and security problems (e.g., trusted platform modules), but the authors do not discuss the problem of unintentional location disclosure [24]. An interesting solution to privacy is given by Lamming et al. [102], that might be adapted to other existing systems like Bluetooth. Their sensing devices use cryptographic keys to change the identifiers of sensor nodes at a fixed time interval. If the key is not known to an attacker, he cannot identify a target beyond this interval.

Intentional location disclosure on the other hand is very useful. One of the most popular questions asked on mobile phones probably is: "Where are you?" On the phone, this question actually leaves a lot of room for negotiation, as Palen and Dourish put it [146]. An automatic system for location disclosure also has to take account for negotiation. Beresford and Stajano propose a system in which location disclosure can be controlled by granularity (e.g., it is not giving the exact location, but only the coarse area) [10]. Elcoate et al. mention access control with role-based rules and spatial, as well as temporal limits [48].

The opposite extreme to surveillance by an authority is what Mann calls "sousveillance" [116]. In this concept, people watch themselves, probably with the same technologies like an authority would do. As a consequence, recordings are done from the perspective of an individual, not "from above," as is usually the case with surveillance. As we will see later, the approach presented in this work is related to the concept of sousveillance, although with the main purpose to learn about oneself and not about others. There might be important potential in these technologies for cooperation and collective action, despite its tendency to panoptic power, as Smith argues [180].

Another topic which we do not want to discuss in detail in this thesis is that of security. Related to Bluetooth technology, viruses can potentially spread from one device to another by this technology [205]. Infected devices pose a high privacy risk to their owners. The Bluediving tool contains a wide selection of attacks on Bluetooth devices that demonstrate its vulnerability [11].

[3] An attacker could do continuous "pages" to a device address [13]. If there is a response, the device is in vicinity.

2.7. Summary

In this chapter the concept of the urban pervasive infrastructure was introduced to understand the system of people, spaces and technologies. The main focus has been to present an overview over the aspects of this system and methods of measuring and analyzing them, as summarized in table 2.2.

UPI	Metrics	Methods
Mobility	Distance traveled Speed Flow Visit duration	Gatecounts Mobile Scanners Emulation Simulation
Temporal structure	Laws and rules Time-based distributions	Inter-connection analysis Longitudinal gatecounts Emulation Simulation
Social structure	Network analysis metrics (e.g. degree, betweenness, closeness)	Longitudinal static snapshots Mobile Scanners Emulation Simulation
Spatial structure	Space syntax metrics (e.g. integration, choice, intelligibility)	Space syntax Simulation
Facts and figures	Statistical characteristics	Wardriving Gatecounts Static snapshots Mobile Scanners

Table 2.2.: Aspects of the UPI and the associated analysis and observation methods

The concepts, metrics and methods presented here may be used to gain an insight into and understanding of the UPI of a city. Such an understanding can have a profound effect on how urban applications are developed and can greatly improve our ability to do so.

Throughout this thesis, we examine different aspects of the UPI. We develop a system to measure the pervasive Bluetooth signals (chapter 5) and apply the method of augmented gatecounts to locations in Bremen, Germany, and San Francisco, US (chapter 7). Further, we present a method to understand the aspect of temporal structure in relation to social structure and spatial structure (chapter 9) and a method to put social structure and spatial structure into relation (chapter 10).

3. Social Context: Concept and Methods

In the previous chapter, we have established the basic frame of reference for our work. We have seen, that cities are evolving into “smart cities,” with sensors that can detect the pulse of the city in real-time and feed it back into the digital counterpart of the city for its users to interpret, when they make their way through the anonymous crowd. This crowd is probably the most defining aspect of a city, as compared to a village, and its internal structure bears rich insight into what makes the city tick as a whole. For an individual, this structure comprises his personal social life, different circles of people, kept safely distinct or interwoven in social complexity.

In this chapter we will try to clarify some aspects of the social fabric the city is made up of. We investigate the concept of social context, both in terms of how it is understood and used in the field of pervasive computing research. To broaden our view on the concept, we will undertake an excursion into the perspectives of other sciences. An examination of tools and methods to analyze, interpret and visualize social context concludes the investigation.

3.1. Context-Aware Computing

With Mark Weiser’s influential work on mobile computing devices, or Tabs and Pads, as he called them, started a couple of new themes of research in computer science [196, 197]. Besides issues of miniaturization, power supply, wireless network access, and new user interfaces for these devices, the usage of the devices in general came into the focus of consideration. Weiser was intrigued by the idea of getting the computer out of the primary focus of attention, to enable the users to fully focus on their tasks and not on the machine [198]. He quickly came to realize, that “This is not a graphical user interface (GUI) problem, but is a property of the whole context of usage of the machine [...]” ([197], p. 76).

It was clear from the beginning, that the new mobile devices could not live in isolation, but had to be integrated into the existing infrastructure of workstations and the Internet. More surprisingly, they were also integrated into office environments in the form of active badges, transmitting the identities of their wearers to nearby receivers by infrared. Thus, the network of receivers and a connected server sensed the location of their users and enabled applications on top of this information [191]. Schilit and Theimer mentioned a couple of such applications, among others “To keep record of located-objects and persons one has encountered” ([164], p. 23), e.g. for later retrieval, and “To detect nearby people” ([164], p. 23), e.g. to trigger reminders or actions based on this information. In another article, Schilit et al. reinforce the meaning of social information in contextual computing systems: “Three important aspects of context are: where you are, who you are with, and what resources are nearby. [... More than location, context includes] even the social situation; e.g. whether you are with your manager or with a co-worker” ([163], p. 85).

3.1.1. Definitions of Context

With the beginning of research on context-aware computing systems, context and context-awareness have been defined in different ways. Often, enumerations were used to describe the concept:

- Location, nearby people, devices in the environment, and changes of these aspects [163]
- Location, proximate people, time, etc. [20]
- Location, time, temperature or user identity [161]
- Physical, social, emotional, and mental (focus-of-attention) environments [39]

Location of the user has been an integral part of context-awareness in most definitions. To build applications taking this part of context into account, various location systems have been designed for this purpose (e.g. [191, 192]).

The social context, including the people in proximity to the user, was also mentioned in many of these definitions [163, 20, 161, 39, 164, 168, 91, 165, 166]. It was usually thought that this could be easily implemented on top of a location system together with an infrastructure calculating the proximity of users. E.g., research around the active badge system was generally based on its server infrastructure, that was necessary to receive the badges’ identification signals. Besides problems with the privacy protection of its users, the system was useless out of reach of the infrastructure. Initially, Lamming and his colleagues used the same system to design an automatic diary [104, 103]. This device was able to record the location of the user and encounters with other people. It used this information as a key to the retrieval of exchanged documents and as a useful memory aid in its own. Later, they advanced the system to cope with the shortcomings of privacy protection and dependence on an infrastructure

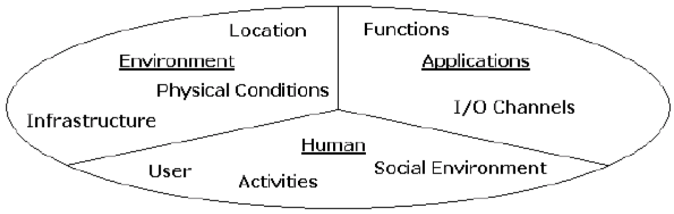


Figure 3.1.: 3D context space [168]

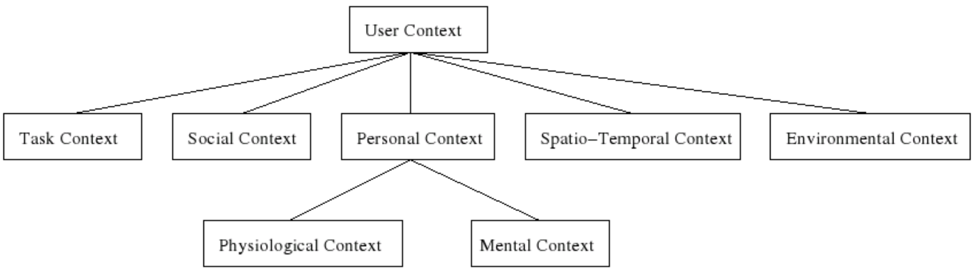


Figure 3.2.: Context hierarchy [91]

by the introduction of peer-sensing capabilities [101, 102]. They developed the SPEC devices that were able to exchange identity information on their own when they were in close proximity of approximately five meters. A server or sensor infrastructure was not necessary. With these devices, they found that proximity information was much easier to obtain than absolute location information, and that it was more important to their research of activity recognition and an automatic diary.

Based on the early context definitions, even broader ones were given. Dey et al. "define context as any information that characterizes a situation related to the interaction between users, applications, and the surrounding environment." ([40], p.100). Chen and Kotz account for the distinction of active and passive contexts in their definition: "Context is the set of environmental states and settings that either determines an application's behavior or in which an application event occurs and is interesting to the user." ([27], p.3).

To get a better picture of what context in pervasive computing is, Sears et al. suggest to understand it as a model with three dimensions: human, environment, and application ([168], figure 3.1). For the design of a system with many different aspects of context-awareness, Kofod-Petersen and Aamodt composed a taxonomy ([91], figure 3.2). Schmidt proposes a similar hierarchic approach, which additionally accounts for changes in time [165].

These hierarchical views of context might suggest a simplistic view of the concept—that we could tackle the whole concept with a divide and conquer strategy. An ethnomethodological study by Tamminem et al. highlights the complex interplay between elements like location, time, task and social context in everyday scenes in urban environments [183]. Any definition of context seems either to be too narrow to include each aspect, or too general to be useful for practical application. As such, we agree with Dourish: “Context is a slippery notion. Perhaps appropriately, it is a concept that keeps to the periphery, and slips away when one attempts to define it.” ([42], p. 29)

In summary, it can be noted that the relevance of the social component of context, including nearby people, was recognized in most definitions. Nevertheless, in most writings the focus is on the most trivial part of social context: the individual identities of surrounding people. A detailed discussion of social context with its separate aspects has not been undertaken. We do not strive for a strict definition—especially after we have seen the difficulties in attempts of related definitions—, but rather need to sharpen our understanding of social context and identify aspects that lend themselves to our analysis. But before we continue to examine pervasive computing examples, we depart from computer science and undertake an excursion into the perspectives of sociology, psychology, geography and economics on social context.

3.2. Aspects of Social Context

Similar to the notion of context in general, social context is a concept that is difficult to grasp and to operationalize. In trying to do so, one naturally runs the risk of either generalizing too much or taking a naive and limiting point of view, as explained above. As we have seen in the last section, pervasive computing literature has the tendency to put social context on the same level with other components like “task context” and the “spatio-temporal context” (see figure 3.2). While this might make sense for the purpose of sensing and data collection, it is a limiting perspective from an application’s point of view.

3.2.1. Mantovani’s Three-Level Model

The psychologist Mantovani developed a model to support HCI (human computer interaction) researchers and practitioners in design of IT tools, by clarifying the context every usage of tools is embedded in. The three-level model describes, how tools are embedded in everyday situations that are in turn embedded in the social context of the users (figure 3.3). Level one is shaped by the interplay of cultural models or social norms (structure) and actions. Social norms are usually valid for a long term, but are nevertheless challenged and changed by actors. This development is denoted by “history.” The background of level one bears individual situations, described by level two, showing how individuals develop their goals. Basically, goals arise from the opportunities one experiences as well as one’s interests. Situations include the presence and

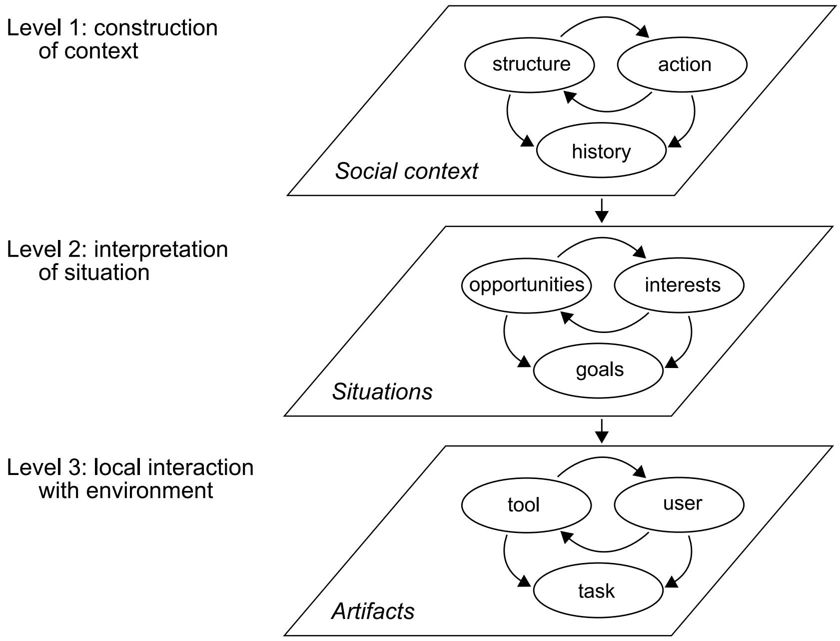


Figure 3.3.: Three-level model of social contexts (adapted from [117])

the actions of other actors that may determine the opportunities a situation offers. To reach a concrete goal, it has to be broken down into separate tasks (level three). A task can then be solved by the usage of tools, such as computers.

Included in this model is the assumption, that social context also lies within people as an essential part of their identities and does not only surround them. Another implication from his model is the subjectivity of social context. Although social norms are shared by the members of a culture, or at least a part thereof is, the interests of individuals and the opportunities they discover are very personal. Thus, social context is subjective and unsharable as a whole.

The three-level model introduced a high level view of social context. In the following sections, we take a deeper look at the concept of milieu (3.2.2), the specific features of urban environments (3.2.3) and the meaning of proximity between individuals (3.2.4).

3.2.2. Milieu

Like the first level in Mantovani's three-level model, the milieu in sociology includes factors determining the development, expansion and modalities of social action. It also describes the factors determining the structure of social entities. As such, it includes the geographical environment, climate, norms, legislature, economy and politics influencing social entities [56].

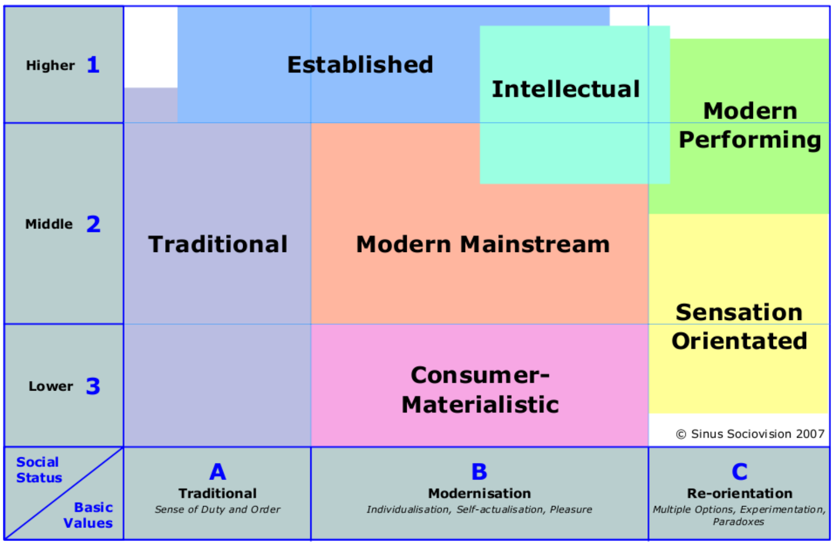


Figure 3.4.: Milieus in Western Europe [175]

At a very coarse level, a milieu might characterize the general culture of a whole country. Milieus can also be used at a finer level to separate different social groups within a society. As an example, Sinus Sociovison developed such a model and offers it for the purpose of product marketing (figure 3.4). Basically, it incorporates the criteria of social status (income/wealth) and basic values (traditional, modernization, re-orientation), to identify target groups for specific marketing campaigns.

According to Durkheim, the most important factors determining the inner milieu—the factors inside a society—are *volume* and *dynamic density* [43]. With volume, he refers to the number of entities a social system is composed of. The dynamic density is a factor, consisting of the number of entities being in relationship to each other, the intensity of interactions, and their moral and normative connectedness. It is interesting to note, that he tries to explain the complex mechanisms within society by a number of relatively simple factors. Factors, that can be observed and quantified, and when combined, result in a picture of the broader social context.

There are many more measures that can be used to find structure in society. One recent example is an analysis by the New York Times [131]. They combined the local popularity measures of a number of movies from the movie rental Netflix with the visualization on a map (see figure 3.5). This method provokes thoughts about how different neighborhoods compare in regards to the values and norms of their residents.

3.2.3. Urban Environment

According to the UN's population database [38], 48.6% of the world's population lived in urban areas in 2005. In the more developed nations, this number was 74.0%. These

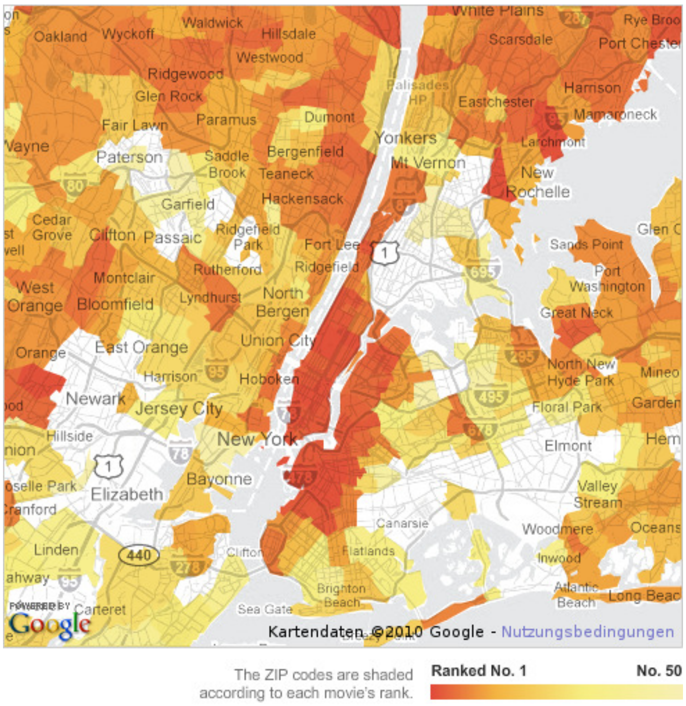


Figure 3.5.: Popularity of the movie "Rachel Getting Married" in different regions of New York [131]

figures are predicted to increase during the next years: to 69.6% for the world's and 86.0% for the more developed nations' population in 2050.

The qualitative differences to rural life were studied by Fischer [52] and Milgram [123] among others. Especially the latter made experiments with a concept important for this thesis and pervasive computing in general. Milgram observed that a new class of relationships has emerged in cities. While in rural areas, relationships could be grouped in either of two classes—*familiar people* and *strangers*—he argued that there was a third one—*familiar strangers*—in cities.

In rural areas, according to Milgram, basically all residents of a village know each other. People from other villages or towns, in contrast, are strangers. In cities, due to the sheer amount of people met during daily routine, urbanites are confronted with cognitive overload. Milgram quantified this overload by comparing the amount of people one can possibly meet within a ten minute radius in three areas in the United States (see table 3.1). Thus, the communication possibilities offered by great cities increase dramatically, but clearly exceed the cognitive capabilities of people. The concept of the familiar stranger explains one form of adaptation to this overload. Milgram's definition of a familiar stranger is that it is person who is encountered repeatedly, but never interacted with. Typically, familiar strangers are encountered on the bus during one's daily way to work or while repeatedly visiting the same recreational facilities.

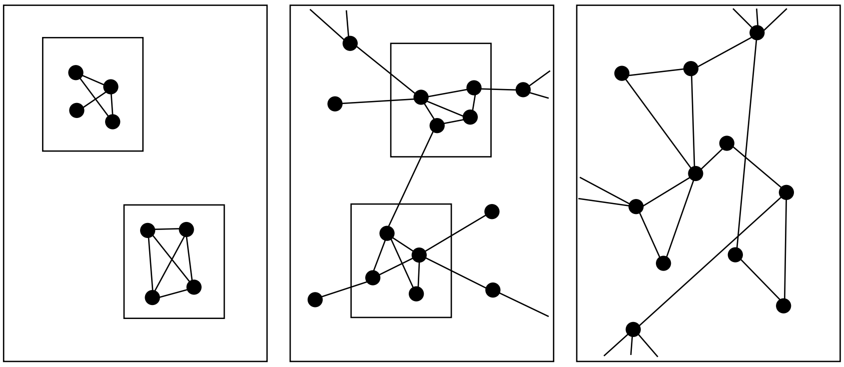
Location	Potential meetings
Nassau County	11,000
Newark	20,000
Manhatten	220,000

Table 3.1.: Comparison of potential meetings within a ten minute radius in rural and urban areas, figures from 1969 [123]

In an experiment, Milgram observed, that people were more willing to help a familiar stranger than a complete stranger, when he was in need of help.

The structure of urban areas also has an impact on the choice of friendship and other associations. Compared to rural areas, relationships are not primarily formed on the basis of physical proximity. Instead, the choice of the neighborhood often has a "status-differentiating component" ([123], p. 48). Urbanites have the possibility to choose from a variety of enterprises to work for, churches to worship in and taverns to socialize at. According to Huckfeldt, it is this complex interplay of choices that determines the social context of individuals: "[...] a person's social context is often seen as being individually unique because it is created from the particular set of environments within which that person resides." ([76], p. 653). Thus, physical proximity is still an important factor in the development of personal relationships with neighborhood being just one component in the whole set of environments. Huckfeldt assumes, that the whole environment "controls the likelihood of encounters within and between social classes, but that individuals exercise discretion in deciding whether to turn an encounter into an association." ([76], p. 654). Another implication of this point of view is, that separate environments, like enterprises, churches and taverns, create their own social contexts, made up of the people being part of them.

Wellman examines the phenomenon of the transformation of society in cities from a different perspective [199]. He uses social networks to explain the shift and also draws a connection to information technology (social network analysis is examined in detail in section 3.4.1). Following his argument, social structure in rural villages before the rise of communication technology can be described by the *little boxes* model (figure 3.6(a)). Communities of individuals were isolated from other communities. Through increasing urbanization and telecommunication, the boundaries became permeable, shifting to what he calls *glocalization* (figure 3.6(b)). In the extreme, this development could lead to a *networked individualization* (figure 3.6(c)), where the once defining boundaries of place (home, workplace) vanish and the individual acts mainly within his own network.



(a) Little boxes (b) Glocalization (c) Networked individualism

Figure 3.6.: From little boxes to networked individualism: three models of community and work social networks (adapted from [199])

This development can not only be observed through patterns of communication and relationships. The public spaces in cities are also shaped by networked individualism. Hebbert observed, that the street was once a place of collective memory, expressing group identity through architecture, symbols, street names and the traces left behind by the patterns of everyday life [65]. Anyhow, he argues, public places nowadays lost their function as collective memory, e.g. due to privatization and shopping malls, and dissipated into cyberspace.

Urbanization, telecommunication and mobility have clearly shaped our society and redefined our usage of space. These developments have not made place irrelevant (at least not yet, and probably never), but for sure they added complexity to the relation between geographically and socially proximate entities. In former rural villages, our home defined our social contacts. In our modern cities, complemented by the world-spanning cyberspace, this situation has become complicated, but still, face-to-face meetings have not lost their power, as we will see in the next section.

3.2.4. Meanings of Spatial Proximity

Tobler's first law of geography states: "everything is related to everything else, but near things are more related than distant things" ([185], p. 236). He used this rule to motivate the application of cellular automata to the simulation of population growth in the Detroit area. Barnes in contrast objects his simplistic approach, stating that human behavior is too complex to be described by law-like statements [8]. He argues, that there is too much messiness and heterogeneity in human behavior—an argument that we must accept, at least when we take a single human into account, who will probably always surprise us with unforeseen actions.

But no matter how much we are convinced that an individual's actions can never be modeled by an algorithm, we inevitably find that sociologists create and validate models showing that groups of people regularly follow such simple rules. In relation to our interest in the connection between spatial collocation and social context, we get support by Huckfeld, who developed a mathematical model describing the influence of the social context of the neighborhood on the choice of friendship [76].

Torre and Rallet [186] take a practical approach to geographical proximity from their perspective of economics. Regarding the organization of large companies with offices in geographical distant places, they value geographical proximity of actors for certain types of interaction—in particular for services or the sharing of knowledge. Geographical proximity leads to knowledge spillovers in firms, and especially in the beginning of new projects, geographical proximity of the team is a necessity. Random meetings are an important factor. Later, they say, can geographical proximity be replaced by organizational proximity and face-to-face meetings are less important.

Another study in an economic context was undertaken by Neff [130]. She examined networking events—from cocktail parties to seminars—in New York's new media industry. One might think, that especially these people, who are responsible for many achievements in cyberspace, would abandon geographical proximity altogether. But Neff learned quite the opposite: networking events "mediate access to crucial resources within the industry" (p. 134), and thus have become more, and not less, important.

Psychologists have found quite simple rules in human behavior, that might provide an explanation to the effects of proximity we have described in various examples. Zajonc discovered, that even the repeated *mere exposure* of material to persons improved their attitude towards it [207]. This effect was demonstrated with abstract material. Simplified Chinese letters were used for this purpose. He repeated the experiment with photographs of unknown persons and could show the same effect. Thus, the theory posits, that attitude improves, regardless of the presented material. Nevertheless, some material might be more attractive overall, compared to other. He found that attitude improves on a log-scale with the number of exposures.

Lee presents empirical evidence, that the mere exposure effect can be explained by *uncertainty reduction*: our positive affect to stimuli increases, because we become familiar with it [106]. Of course, there is a multitude of considerations, including *tedium*, which causes positive attribution to decrease with boredom in the long run.

These findings might present an explanation for the observations of Milgram [123] regarding the improved affect towards *familiar strangers*. We are exposed to them repeatedly, thus reducing our uncertainty and improving our affect towards them. Figure 3.7 shows a simplified model of proximity and the development of relationships, that summarizes this discussion.

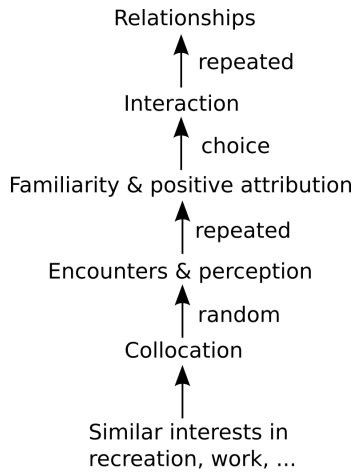


Figure 3.7.: Model of the development of relationships from similar interests

Distance	Appropriate Relationships and Activities	Sensory Qualities
Intimate Distance (0 to 0.5 meters)	Intimate contacts, e.g. making love, comforting) and physical sports (e.g. wrestling)	Intense awareness of sensory inputs (e.g. small, radiant heat) from other person: touch overtakes vocalization as primary mode of communication.
Personal Distance (0.5 to 1.2 meters)	Contacts between close friends, as well as everyday interactions with acquaintances.	Less awareness of sensory inputs than intimate distance; vision is normal and provides detailed feedback; verbal channels account for more communication than touch
Social Distance (1.2 to 3.7 meters)	Impersonal and businesslike contacts	Sensory inputs minimal; information provided by visual channels less detailed than in personal distance; normal voice level (audible at 6 meters) maintained; touch not possible.
Public Distance (more than 3.7 meters)	Formal contacts between an individual (e.g. actor, politician) and the public	No sensory inputs; no detailed visual input; exaggerated nonverbal behaviors employed to supplement verbal communication, since subtle shades of meaning are lost at this distance.

Table 3.2.: Edward T. Hall's spatial zones [62]

When we take a look at the relationship between geographical proximity and social relationships from a different perspective, we find that the physical distance between two persons tells a lot about their relationship as well as cultural norms. The anthropologist Edward T. Hall introduced the concept of *proxemics* to measure this relationship and to study the social uses of space [62]. In his studies, he found that relationships and activities can be classified based on the spatial distance between two persons. He identified four different classes of distances, as given in table 3.2: intimate (0m – 0.5m), personal (0.5m – 1.2m), social (1.2m – 3.7m) and public (more than 3.7m). Thus, by only observing the distances between persons, their relationship can be inferred roughly. Formal contacts can be distinguished from impersonal and businesslike ones, contacts between close friends from intimate ones. Further, he argues that the influence of two persons loosely is inversely proportional to the square of their distance. Nevertheless, his classification is not universal: the distances he gives are dependent on cultural norms. While the given values are valid for US Americans, they may differ for other cultural areas, such as Europe or Asia. There are also environmental factors, that might cause altered distances, e.g. extreme background noise might force people to get closer to have a conversation. As two example actions at intimate distance—comforting and wrestling—demonstrate, there is ambiguity involved in the mapping of observed distance to relationship.

3.3. Applications of Social Context

Social proximity is ambiguous, but nevertheless meaningful to people's relationships. As such, it is seen as a basic building block of social context and applications that build on social fabric, as we have pointed out in the beginning of this chapter. It comes to no surprise that social proximity is demonstrated in a range of research papers in pervasive computing. But there are as well concepts to exploit proximity and mediate between two persons, identify common goals and encourage interaction. Creation of an awareness for a whole group of people, as well as analysis of the surrounding milieu, can also be found in research. In the following, we review pervasive computing research for applications of social context.

3.3.1. Memory Augmentation

The first wave of pervasive computing research was conducted at Xerox Research, where a range of novel mobile computing devices—pads and tabs—together with wired infrastructure for network connectivity and location awareness provided the infrastructure. Lamming and Flynn used location awareness to infer proximity of devices, and thus of its users [103]. Figure 3.8 shows a diary application running on a

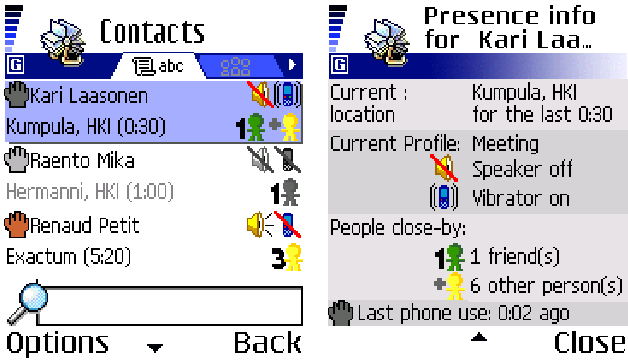
Figure 3.8.: Forget-Me-Not on PARCTAB mobile computer [103]

tab. Different people met and locations visited are displayed by small icons, organized in a chronological manner to resemble a diary. Its purpose was to cue memory recalls of the user—like a memory prosthesis. Social context, the people around us, are a strong cue for our episodic memory. The idea was continuously developed further with more sensing capabilities and independence from the office infrastructure [101].

Kern et al. describe a system, that can automatically annotate the video and audio recordings of a meeting to facilitate appropriate retrieval of specific episodes [87]. Audio and acceleration sensors are used to sense the people in a meeting, determine who is talking, detecting the shaking of hands. A study conducted at a conference with the iBand electronic bracelet focused on social networking [86]. The bracelet detected shaking of hands with other iBand users and recorded their identities. Later, all contacts could be retrieved, which served as a good memory assistance for the users.

3.3.2. Interpersonal Awareness

A direct application of social context is the augmentation of interpersonal awareness, reaching out beyond our innate senses. Social beings as we are, we want to feel connected to our friends and families. Interpersonal awareness devices—or IPADs, how Holmquist calls them—can facilitate this in a number of ways. His Hummingbird is one instance, which translates the distance to other Hummingbird users into a humming sound, getting louder the closer the other person comes, in a 100m radius [72]. The devices do not rely on any infrastructure, so that their usage could not only be studied in the office, but also on a rock festival or on a skiing course [195]. They found the devices very useful for the organization of social activities, e.g. finding colleagues and friends and meeting for lunch.



(a) Contacts overview (b) Contact details

Figure 3.9.: The context contacts application for Symbian [156]

Smith et al. studied, how people were making use of automatic location disclosure in their daily life [179]. They used mobile phones with Cell-ID positioning to determine the rough location of their probands. The probands could set their own rules of who had access to their location. The study shows, that this automatic disclosure of location—one person can query for the location of another without confirmation—is mostly used and appreciated by spouses.

Context lends itself for integration into common mobile phone programs, as demonstrated by Raento et at. [156]. They extended the contact list of a Symbian phone to include the location of a person, as well as their social context—proximate friends and other people (see figure 3.9).

A similar system was conceptually designed by Oloffson et al. in another setting: they focused their study on the use of technology by the visitors of a music festival [144]. In an ethnographic study, they found that SMS was commonly used to connect to friends and arrange a meeting with them on the festival. However, they found that this method is unsuitable. Instead, they designed a system based on GPS-enabled mobile phones, which indicated the locations of friends in relation to the own location on a map of the area.

3.3.3. Encouraging Interaction

While interpersonal awareness devices provide contextual information for familiar people, the technology can also be used to encourage interaction between strangers. The Lovegety is a commercially successful example, that signals a match in interests [81] (see figure 3.10). There are separate devices for males and females. Users can select one of a limited set of interests (talk, karaoke and get2). Whenever a device of

Figure 3.10.: Male and female Lovegety devices (photograph by [203])

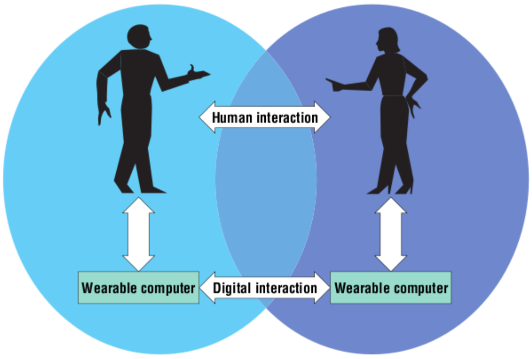


Figure 3.11.: Wearable communities to enhance cooperation [92]

opposite sex with the same interest is in proximity of five meters, both devices flash and beep, thus encouraging the couple to interact.

More sophisticated dating applications were described by Eagle and Pentland [45] and Beale [9]. Their systems are based on mobile phones. Discovery of peer devices is accomplished by Bluetooth. Databases with personal profiles enable greater detail in specifying the partner to look for. Esbjörnsson et al. designed and evaluated a similar system for motorcyclists [51]. Their ethnographic study highlights, that bikers are interested in sharing information, such as routes, and in socializing. Thus they welcome a device to encourage interaction between them.

The above examples all used a kind of matching of interests and profiles to match users. Terry et al. in contrast describe a system, which infers a match from social structure [184]. People are introduced on the basis of mutually common acquaintances.

3.3.4. Supporting Cooperation

A further class of pervasive applications of social context comprises examples with the purpose to support cooperation between individuals. Kortuem and Segall outline a system of wearable computers that enriches and supports face-to-face interactions (see figure 3.11, [92]). The Genie application, which runs on top of this infrastructure, helps its users (students on a campus) to find solutions to any problem. They enter a set of questions into their wearables. Whenever two Genie users encounter each other, the questions of one person are displayed to the other. If he knows the answer, or would like to discuss the topic, the two can start a face-to-face conversation. In a similar way, Kortuem and Segal's WALID application supports cooperation [92]. With WALID, users can organize their todo-lists (e.g. for shopping). When two users meet, their computers look for opportunities of optimization, probably resulting in an exchange of tasks, so that each person can finish faster.

These examples are based on real-world meetings, and not only on cyberspace, which results in a more trusted relationship and thus fosters cooperation. Schneider et al. designed a system that could be used to reinforce trust in such settings with a decentralized reputation framework [167], which is important if cooperation with strangers is desired.

The name tags of Borovoy et al. are a simple example for cooperation [17]. They are initiated with a set of answers to a number of common questions. Whenever two users of these name tags face each other, the tags indicate the number of matching answers. These tags are a tool for cooperation because they help like-minded people to find each other. Kikin-Gil developed a concept for sharing planned activities and experiences for groups of friends to increase their social effectiveness [88].

An economic and environmentally sound approach to encourage ride sharing in urban environments is taken by Xing et al. [204]. By matching routes and positions of vehicles and pedestrians, pick up and transit points can be calculated, and pedestrians could reach their destinations easier and faster, compared to using public transport. Seitz et al. investigate a similar scenario and focus on algorithms to form local groups [169].

Cooperation can also occur implicitly, hidden from the attention of the users of the technology, but working for them nevertheless. This is for example the case in delay tolerant networking scenarios. Human beings become carriers for data packets that silently jump on and off their mobile devices [36]. To efficiently forward packets through ad-hoc networks, Hui et al. studied the mobility and meeting patterns of conference visitors [77].

3.3.5. Group Awareness

Social context can also be leveraged to create awareness within a group of people for the activities that take place in the group. This is similar to interpersonal awareness, but extends to a group of people. A number of systems for group awareness work

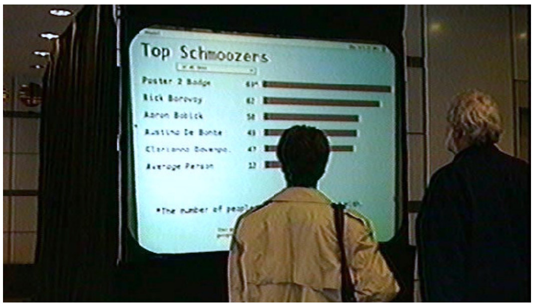
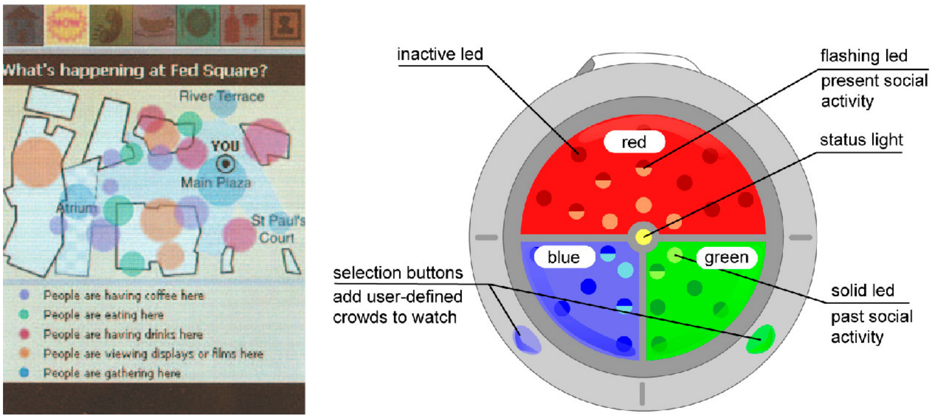


Figure 3.12.: Community mirror to display interaction data and meme flow [18]



(a) Display of social real-time activity on a mobile phone [90]

(b) Jabberwocky device to visualize familiar strangers [149]

Figure 3.13.: Two examples of milieu awareness

on the scale of a room or a building, and the group of people using that facilities. Conferences are also popular scenarios for group awareness systems.

The MusicFX system can automatically determine a common taste of music for the visitors of a fitness center [119]. Based on lists of music the present group members like or dislike, it plays a selection for everybody's pleasure. Sawhney et al. designed a mechanism to select news content for a public display based on the people around [162].

A variation on the name tag application by Borovoy et al. mentioned in the last section is called Meme Tags [18]. The wearers of the tags can enter a short meme, which is displayed to face-to-face contacts. Memes can be picked up and carried on. A big screen in the venue projects currently hot memes for everybody to see, creating a summary of the memes of the whole group (see figure 3.12).

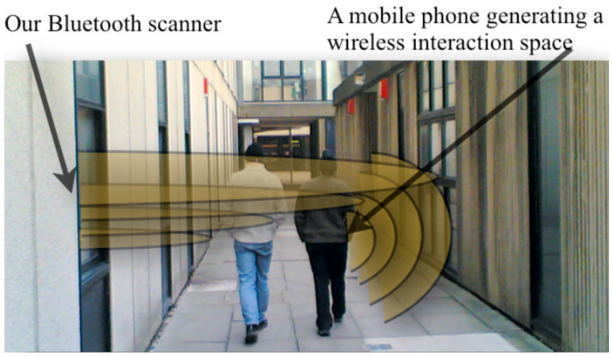


Figure 3.14.: A Bluetooth gate [145]

3.3.6. Milieu Awareness

Applications that extend from closed groups to the population as a whole, including potentially everybody in a particular area, are measuring parameters of the milieu. Kjeldskov et al. describe the so-called Just-for-Us system in Melbourne, Australia, spanning the whole of Federation square [90]. The system senses the locations, quantity and activities of people, aggregates the information and visualizes it on a mobile phone (see figure 3.13(a)).

While Just-for-Us takes an allocentric perspective with a number of stationary scanners to sense activity in real-time, the Jabberwocky device measures and displays the milieu from an egocentric point of view [149]. Following the observation of Milgram, that familiar strangers give us some degree of confidence (compare section 3.2.3), they use periodic Bluetooth signals from mobile phones to indicate the comfort a specific place affords (see figure 3.13(b)).

Nishimoto et al. sketch an infrastructure of personal, mobile musical instruments and players, opportunistically connected to stationary kiosks [140]. In this system, tunes are transmitted between mobile devices when in proximity and elaborated upon by their users. The stationary kiosks pick up and send out tunes, that are popular in the surroundings, and are thus a reflection of the regional musical culture.

3.3.7. Measurement and Analysis of Human Behavior

The RealityMining experiment was a large-scale study to automatically measure and analyze human behavior via mobile phones [46]. About one hundred students and faculty staff members were monitored over the course of nine months to gather a rich dataset. A number of approaches to analyze the data were proposed, among others the eigenbehavior method with the goal to characterize social behavior.

While the RealityMining data was recorded from mobile scanners, the Cityware study was based on a number of stationary scanners, installed at specific places throughout the city of Bath, UK [145]. These Bluetooth scanners created a number of gates, automatically counting and identifying the passers-by, and giving rise to realistic mobility models (see figure 3.14).

3.3.8. Summary

The examination of social context in pervasive computing has shown us a broad scale of applications, ranging from the augmentation of the user's memory, through awareness applications for individuals, groups and whole populations, to projects with the goal to measure and analyze human behavior. A selection of concrete examples has demonstrated the basic properties of the different groups of applications.

The classes presented here make no claim to be complete. There are also different works that leverage social context for privacy control [48], or which are concerned with specialized sensors (e.g. the inference of social context from ambient sounds [178]). Rather, these classes were chosen to correspond to the notion of social context, we gained in section 3.2 (probably with the exception of the memory augmentation applications). In comparison, we can see that we find many analogies. When we look at the concept of social context from the perspective of pervasive computing, we find that it is centered around sensing the identities and quantities of people in the proximity to a person or a place.

3.4. Tools and Methods for Analysis and Visualization of Social Context

Most applications of social context we have reviewed have no need for sophisticated models. Social proximity information is consumed straight forward, often in combination with some kinds of profiles to match. Nevertheless, the applications for milieu awareness (section 3.3.6) and the measurement and analysis of human behavior (section 3.3.7) have introduced new levels of complexity, both in terms of the sheer amount of data and the necessity to extract meaning. In the following, we introduce a number of advanced techniques to understand and analyze such kinds of data. Our special focus is on visualization techniques.

3.4.1. Social Network Analysis

The primary assumption of social network analysis is, that the observation of relations between separate entities is crucial to the understanding of the behavior of a social system as a whole. From this perspective, the observation of properties of actors are

	Ana	Bob	Chris	Dick	Elena
Ana	-	2	1	1	0
Bob	2	-	0	1	0
Chris	1	0	-	0	3
Dick	1	1	0	-	0
Elena	0	0	3	0	-

Table 3.3.: Example sociomatrix describing meetings between five actors

not enough to characterize behavior in most cases, instead the interplay between the actors is more revealing.

Social network analysis considers social structure from the bottom up. Social structure is modeled by ties between single actors, which are composed into larger units of groups, which are in turn composed into even larger units of networks [193]. In the whole model, relational ties between actors are primary, while the attributes of actors are secondary. On top of the relational data, social network analysis incorporates a number of mathematical methods to aggregate and compute various properties.

Modern social network analysis has its roots in Moreno's sociometry of the 1930s [128]. Since then, it has been extended and applied to various studies. Further extension can be attributed to the recent development of computer programs for convenient data management, application of mathematical methods, visualization and exploration of network data, e.g. UCINET [16] and Pajek [142].

Fundamental Concepts We would like to give a simple example to explain basic properties of social network data. Table 3.3 shows relational data about five actors. Network data is often represented as a matrix. The separate cells contain data about a specific relational variable. Here, meetings between the five persons are represented. A variable might be directed or undirected. This example shows an undirected variable, resulting in a symmetric matrix. A variable can be categorical (e.g. discrete or unordered values) or valued to describe various relationships, e.g. intensities, distances, similarities or—like in the example—repetitions of meetings.

There is a large variety of relationships which can be represented this way, e.g.:

- evaluation (e.g. rating),
- transfer of resources (e.g. money, email),
- association or affiliation (e.g. to groups, places, events),
- movement between places or statuses,

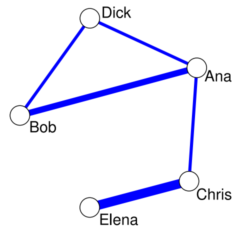


Figure 3.15.: Example social network

- physical connection (e.g. road, river),
- formal roles and
- kinship.

For this thesis behavioral interaction is of special interest: "Interactions involve the physical interaction of actors or their presence in the same place at the same time. Examples of interactions include: sitting next to each other, attending the same party, visiting a person's home, hitting, hugging, disciplining, conversing, and so on." ([193], p. 38)

Visually, a social network can be represented as a graph (figure 3.15 visualizes the data of table 3.3). The respective layout is not directly derived from the data, but depends on the author's intention. Different layouts may clarify different interrelations in the data.

It is possible to apply various concepts well-known from graph theory to a social network graph. A graph can be partitioned into a discrete set of disconnected *components*. *Isolates* are components of size one. In the example, there is only one component $N = \{Ana, Bob, Chris, Dick, Elena\}$ and no isolate.

To characterize the internal structure of connected components, the shortest path between any two nodes can be measured, denoted *geodesic*. The length of such a shortest path is called the *geodesic distance* $d(i, j)$ of two nodes i and j. The *characteristic path length* of a component is defined as the average geodesic distance of all pairs of nodes.

The *clustering coefficient* measures the cliquishness of a network and is defined as follows. Let k_n be the number of neighbors of a node n. Then, there are at most $m_n = k_n(k_n - 1)/2$ possible connections between the neighbors. With $C_n = k_n/m_n$, the clustering coefficient is defined as the average of C_n over all n.

Two-Mode Networks The network presented above is a one-mode network. It is composed entirely of actors of the same type. In contrast, a *two-mode* network

	University	Lake
Ana	1	0
Bob	1	0
Chris	1	1
Dick	1	1
Elena	0	1

Table 3.4.: Example two-mode affiliation matrix, showing whether a person visits a certain location (1), or not (0)

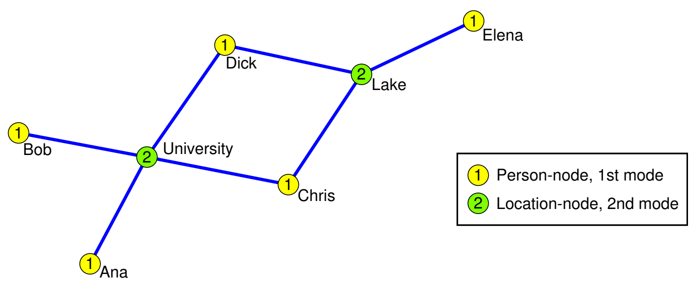


Figure 3.16.: Example two-mode social network

represents relations between different types. A common example of a two-mode network is an affiliation network, showing the relations between actors and events (or companies, etc.). In table 3.4 and figure 3.16 an example is given, which shows relations between persons and locations. A two-mode network is represented by a matrix $A = (a_{ij})_{i=1 \ldots n, j=1 \ldots m}$, with n number of entities in mode one (e.g. 5 for the persons Ana, Bob, ...) and m in the second mode (e.g. 2 for the locations university and lake).

Sociologists are interested in such kinds of networks showing co-membership in organizations or co-attendance in events, because they argue, that "individuals' affiliations with events provide direct linkages between the actors and/or between the events." ([193], p.196). As such, affiliation networks do not only tell about the relations between actors and events, but also between the actors themselves, since "joint participation in events not only provides the opportunity for actors to interact, but also increases the probability that direct pairwise ties (such as acquaintanceship) will develop" ([193], p.293). Important properties of these networks are the rate of participation as well as the size of events.

To interpret an affiliation network as a network between the actors, the two-mode network can be transformed to a one-mode network. A common transformation is to count the number of common locations (for the example), visited by each pair of actors. Mathematically, this operation is accomplished by a matrix multiplication of the two-mode matrix A with its transpose A':

$$X^N = AA' \tag{3.1}$$

The number of people a location has in common can be calculated with a similar operation:

$$X^M = A'A \tag{3.2}$$

Bipartite Networks Conceptually, a *bipartite network* is the same as a two-mode network: there are two different sets of entities, with no connections within a set. Only connections between entities in different sets are allowed.

Bipartite networks differ from two-mode networks in the matrix representation and the arising consequences. If we interpret the network in figure 3.16 as a bipartite network, the adjacency matrix is of the form $A = (a_{ij})_{i=1...n+m,j=1...n+m}$, with n and m being the number of entities in each set.

Ego Networks An *ego network* of a node is a special case of a network. It contains the ego node, its neighbors and the connections between both of them. In some cases, the ego node might be omitted to only show the neighborhood [142]. So, the main difference to other networks lies in the population included in the analysis. If the internal structure of an organization is to be analyzed, the population might include all people being part of it. For the purpose of measuring the social context of a person (ego), we might rather include all persons, that ego is in contact with. Most social network analysis methods can be used on ego networks, too, but their interpretation might be slightly different.

Social Position There are several methods to analyze a network and find general structure within. The later calculations in this thesis are based on the concept of *social position*, which will be introduced in the following section. For an extensive review of the topic, see [193].

The idea of the positional analysis is to find collections of actors who are similar in their ties with others. E.g., in a school we might find that all pupils of a class have very similar relations to their teacher, thus grouping them in the same social position. However, pupils in different classes with different teachers do not share the same position. The general analysis of such *social roles* (e.g. teacher and pupil) is another subject not discussed here.

A number of separate steps is required for a positional analysis. Based on relational data, these are:

- measuring the similarity of any two actors in a network,
- partitioning actors into groups based on their similarity and
- representing the positions in a model.

As a measure of similarity, *structural equivalence* is usually applied to group actors into social positions. Lorrain and White defined this notion as follows: if two actors have identical ties to and from all other actors in the network, they are structurally equivalent [110]. More formally, two actors i and j are structurally equivalent, if $x_{ik} = x_{jk}$ and $x_{ki} = x_{kj}$ for $k = 1, 2, ..., g$.

Structural equivalence can be measured by several means, e.g. euclidean distance [21] or Pearson correlation among others. Since the latter is better suited to find a similar pattern, instead of identity in ties, and especially to analyze interaction frequencies ([193], p. 375), it will be used for the later analysis of social positions in this thesis. Pearson correlation for undirected sociomatrices can be calculated with the formula:

$$r_{ij} = \frac{\sum_{k=1}^{g} (x_{ki} - \bar{x}_{\bullet i})(x_{kj} - \bar{x}_{\bullet j})}{\sqrt{\sum_{k=1}^{g} (x_{ki} - \bar{x}_{\bullet i})^2} \sqrt{\sum_{k=1}^{g} (x_{kj} - \bar{x}_{\bullet j})^2}} \tag{3.3}$$

for $i \neq k$, $j \neq k$. $\bar{x}_{\bullet i}$ denotes the mean of the values in column i. These correlations are arranged in a $g \times g$ correlation matrix C_1 with the (i, j)th element equal to r_{ij}. Diagonal elements are excluded from calculation.

Based on the correlation matrix, the actors can be grouped into different positions by *hierarchical clustering* (or other methods, e.g. CONCOR [193]). This widely applied method gives a dendrogram to describe similarity. The similarity classes can then be chosen, e.g. based on the expected number of classes, as suggested by the theoretical foundation.

The last step of a positional analysis is to abstract from the separate actors and the ties between them. Instead, a model is formed to show the more general positions and the ties between. Such a model can be represented as a *reduced graph* or *blockmodel*. With the classes identified in the last step, a blockmodel is formed by

1. permuting the rows and columns of the original sociomatrix according to the classes, so that actors who are assigned to the same position are adjacent in the matrix. As a result, we obtain submatrices characterizing the relations between the positions.

2. summarizing each submatrix to a single value that is characteristic for the whole relation between all the actors of the positions. If the data is dichotomous, the *density* might be a good choice (number of 1's in relation to 0's in the subma-

trix). For valued data the *mean value criterion* is a common approach, where the average value is calculated.

The positional analysis, as briefly presented here, is used extensively for the later analysis of our data in chapters 9 and 10. Overall, this kind of analysis is a broad topic with many more possibilities to apply different methods for different situations. A broad overview would go beyond the scope of this thesis, the respective chapters of [193] give more details.

3.4.2. Visualization of Social Networks

With the conversion of a sociomatrix into a set of nodes with lines or arrows between them, visually graspable versions of the same data are created. While the matrix representation is concise and unique, the graph representation can take many forms and is usually chosen to support a specific thesis and might also be subject to artistic considerations. While the layout of small networks is usually straight-forward (e.g. figure 3.16, p. 52), the layout of large networks is not trivial. Looking at figure 2.2, p. 24, we cannot see any structure within the network, except, that most nodes are connected in a rather fuzzy way, while a few are isolated. Although this impression might be true, the seemingly connected component could also consist of a number of smaller, disconnected ones.

With modern computer programs (e.g. UCINET [16], Pajek [142], MAGE [112] or MOVIEMOL [129]), the visualization of social network data can be used to explore the data's internal structure and search for certain attributes. Therefore, the artistic aspects can be neglected, instead we are interested in a visualization, that is faithful to the structure of the data. Methods can generally be divided into two groups: multidimensional scaling (MDS) and singular value decomposition (SVD) [54].

MDS represents a number of different algorithms and is based on the concept of searching for an optimal solution. Thus, different starting situations might give different solutions, as a local optimum is found. Two popular methods modeling embedded springs are given by Kamada and Kawai [85] and Fruchterman and Reingold [55].

SVD in contrast produces deterministic solutions. A multidimensional space is transformed into a lower dimensional space (typically one, two or three for visualization). These new dimensions are chosen to express the main variance of the data. Different preprocessing methods can be used to emphasize various aspects. Although computers have accelerated and simplified the access and ease of use, the general idea has been used since the 50s [15].

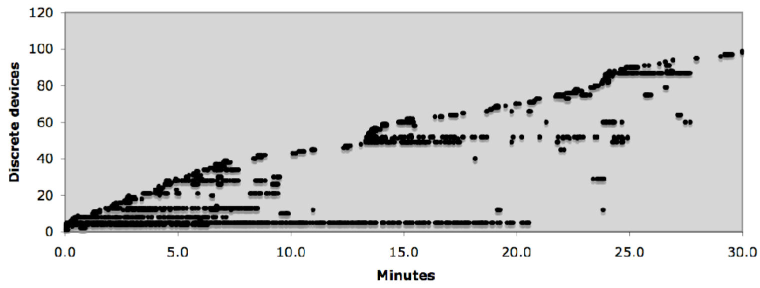


Figure 3.17.: Visualization of Bluetooth gatecount records [145]

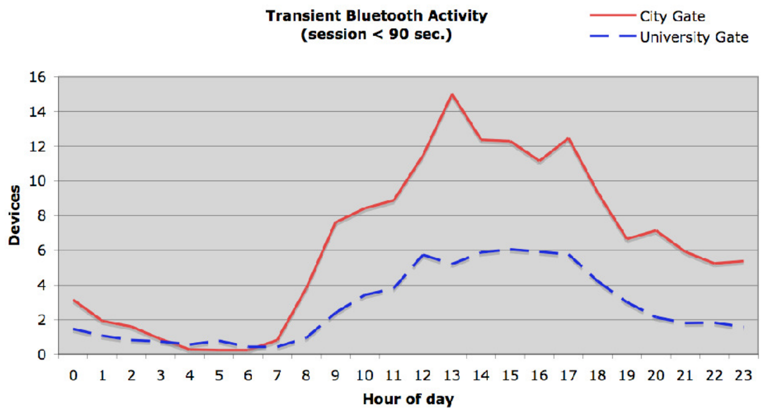


Figure 3.18.: Transient Bluetooth devices for gates on a campus and in a city [145]

3.4.3. Augmented Gatecounts

The observation method of augmented gatecounts was introduced in section 2.5.1. Appropriate visualizations of such data can help to quickly reveal special patterns and characteristics and to compare different datasets with each other. O'Neill et al. plot gatecount data as shown in figure 3.17. Each unique Bluetooth device has a horizontal time-line in the graph. Devices are sorted by their time of first appearance from bottom to top. As a result, the diagonal from the origin at the bottom left to the upper right visualizes devices passing the gate for the first time. The slope and shape are indicators for the rate of new devices. The area below this diagonal indicates repeated or persistent activity of devices already discovered before.

Data of different gates can also be compared by summing up the number of distinct devices discovered each hour (see figure 3.18, only transient devices appearing for less than 90 seconds are included). Such graphs are indicative of the amount of passing devices per hour of the day and show peak times of high traffic.

Further opportunities for analysis and visualization of gatecount data arise, when it is transformed into a network. The nodes in the network are supposed to be the unique

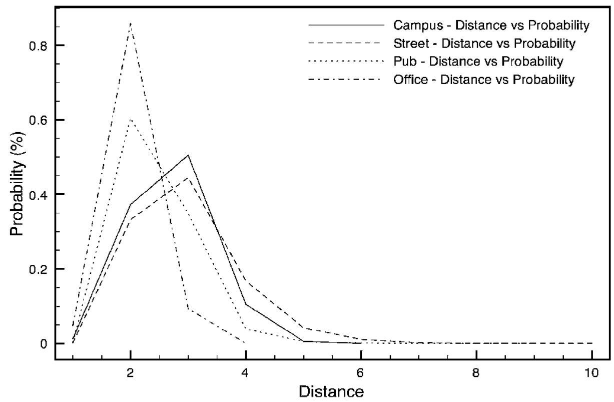


Figure 3.19.: Probability distribution of distance between any two nodes in the network graphs of four scanning sites [95]

Bluetooth devices. To determine the connectivity of nodes, Kostakos and O'Neill look for concurrent appearances of Bluetooth devices in the gatecount data [95]. Two nodes are connected, if they appear together (in a certain time window) at a gate. Figure 2.2, p. 24, shows the resulting network for one gate. Although the network picture might look confusing, this representation enables a number of analyses known from social network analysis (see section 3.4.1). E.g., the geodesic distances in the network can be analyzed and used to compare different gates with each other, as shown in figure 3.19.

3.4.4. Augmented Interpersonal Encounters

The augmented gatecount method described in the previous section is carried out from the perspective of specific places. The RealityMining study, as briefly introduced in section 3.3.7, leveraged the same Bluetooth technology to gather social proximity data from the perspectives of one hundred probands in the study [44, 46]. As such, a number of different analysis and visualization methods were developed for the interpersonal encounters recorded in the study.

Encounters in the RealityMining study are measured differently, than they are in augmented gatecounts. Since each person in the study is carrying a mobile scanning device, an encounter can be determined as soon as a scanner detects another device. Concurrency of two mobile devices at the same time and place is implicit in this consideration. However, the method to detect encounters from gatecounts could be used on augmented interpersonal encounters to find and eliminate measurement errors.

Eagle and Pentland make use of heat maps to show data collected over a long period of time in a compact diagram. Figure 3.20 shows the number of Bluetooth encounters of one person for the duration of a whole month. They further set out to calculate the

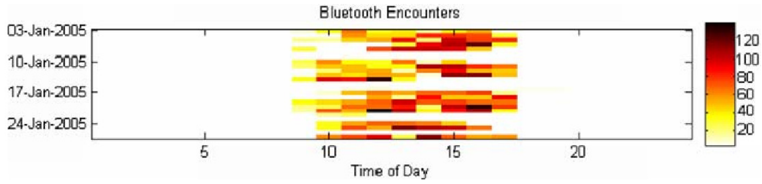


Figure 3.20.: Daily distribution of one person's Bluetooth encounters for one month [46]

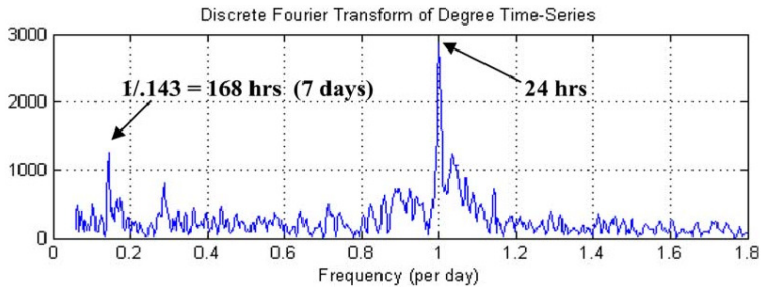


Figure 3.21.: Proximity time-series and organizational rhythms, fourier transformation [46]

entropy of this data, combined with location measured by Cell-ID, to estimate how predictable a person is and how well models of personal behavior can perform.

Further, they show that it is possible to identify rhythms in the data, recurring patterns in time. Therefore, Bluetooth encounters over an extended period of time are summarized similar to figure 3.18. A discrete Fourier transformation of the time-series data reveals prevalent frequencies in the data, e.g. the 24 hours and seven days rhythms (see figure 3.21).

Eagle also investigated the extraction and characterization of separate "behaviors" for individuals or groups of people [44]. As a basis, he used heat maps of Bluetooth activity, as given in figure 3.20. His patterns describe the typical behavior of a subject or group of subjects during a day in terms of the number of Bluetooth contacts over this frame of time. These patterns are explanatory of the greatest variance in the data, thus giving the most prevalent behaviors. Mathematically, this so-called *eigenbehavior* analysis is accomplished by a principal component analysis. Figure 3.22 shows the top three composite behaviors for three different groups of students in the study. E.g., the top left behavior shows the prevalence of the 10:30 coffee break for the business students, the #2 eigenbehaviors indicate, that the incoming lab students have a tendency to go out in the evening compared to the others.

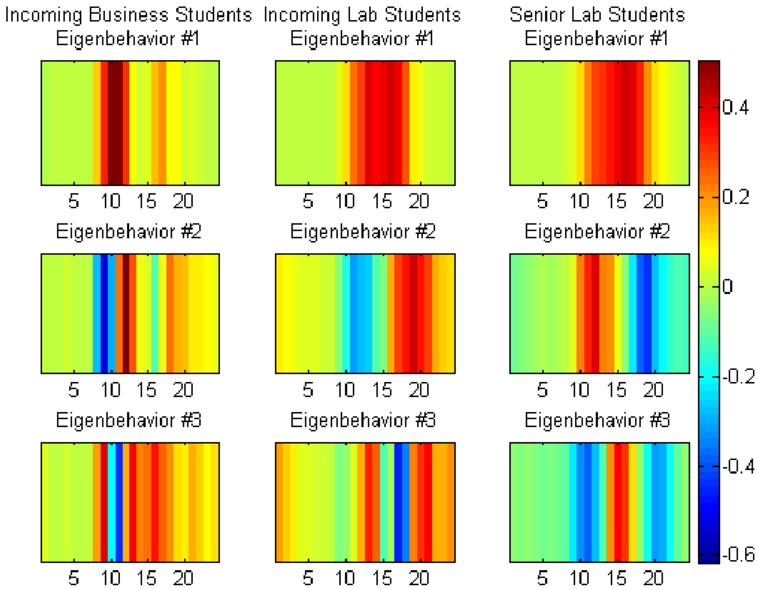


Figure 3.22.: Top three eigenbehaviors for each group of students in the RealityMining study [44]

3.4.5. Encounters

When comparing the notion of encounter in sections 3.4.3 and 3.4.4, we see that the concept is used differently. With the mobile scanners of RealityMining, an encounter happens, when the scanner of one person detects the presence of another one. With the gatecount scanners, an encounter is registered when two devices are detected at the same time at the same place. To have a consistent view of this basic concept, we will propose a definition that takes both perspectives into respect and defines encounters in two different grades: 1st grade encounters are based on experience, 2nd grade encounters are based on observation.

Figure 3.23 illustrates the problem. There are three devices, each with an associated detection radius. A and B, as well as A and C, can detect each other and can thus register mutual 1st grade encounters. A is in the unique position to detect the presence of both B and C and thus registers a 2nd grade encounter between B and C. As we can see, the distinction between both kinds of encounters has consequences for the range of detection. Furthermore, 1st grade encounters require a relatively large number of scanning devices, which was the case in RealityMining. The study in Bath in contrast separated devices conducting the scanning (the augmented gatecount devices) and devices being detected (the phones people were carrying). Also, mixed scenarios are possible, with mobile scanners recording 2nd grade encounters.

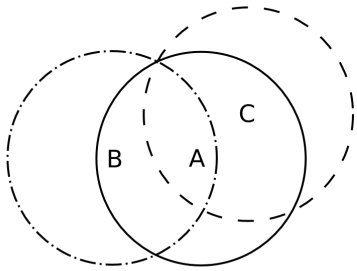


Figure 3.23.: 2nd grade encounter between B and C observed by A

3.5. Summary

Social context is a broad concept, relevant to a number of different scientific disciplines, each with its unique perspective on the concept. We have looked at social context under the premise of urban environment, the overarching scenario for the study to be conducted here.

Most applications we have reviewed, do not reflect the inherent complexities, but rather manage to make good use of the simpler aspects of social context. The applications of interpersonal awareness, e.g., measure social proximity and have developed several methods—from a humming sound to augmentation of a phone's contact application—to present this data to its users.

Recently, we have seen a trend towards the measurement of movement and behavior of large groups of people, both from a stationary perspective, as done by augmented gatecounts, and from an egocentric perspective, as done in the RealityMining study. These large scale experiments with their vast quantity of social proximity data collections, give rise to review well established methods of social network analysis and to find new methods to visually explore and extract meaning from the data.

4. Technologies for Proximity Detection

As we learned in the previous chapter, spatial proximity between persons is an important aspect of social context. In this chapter, we take a detailed look at the different technologies available to build social proximity systems. Based on a review of social proximity systems in research as well as in commercial products, the underlying technologies—ranging from RF-based systems like GPS and Cell-ID to optical systems—are analyzed.

Bluetooth, with its device inquiry method, turns out to be the best choice for our study. The technical specification of this short range communication protocol, now omnipresent in mobile phones and other devices around the world, is discussed. We elaborate the advantages and disadvantages of Bluetooth for our application.

4.1. Approaches to Social Proximity Detection

Social proximity systems can be understood as taking the spatial relationship between two or more persons into account as a basis for various social applications. Such applications may cover areas, such as

- coordination of staff in large organizations,
- generation of an automatic diary for its user,
- visualization of a user's relationships and participation in communities,
- conference networking support,
- interpersonal awareness in general,
- avalanche rescue,
- identification of people by matching profiles, e.g. for general socializing, dating or business introductions,
- or more generally, to understand the urban pervasive infrastructure itself.

Location systems are closely related to proximity systems. In comparison to proximity systems, location systems determine the absolute geographical position of a device. Often, they can be adapted to provide proximity, too. This can be accomplished, e.g. by sending the absolute positions of two users to a server, that calculates the distance between them. Hightower et al. give a comprehensive overview over a number of location systems [66].

Thus, there is a basic architectural difference between proximity systems relying on a server to calculate proximity and systems that can measure proximity directly, e.g. by peer-sensing. Usually, the range of peer-sensing is limited by the used technology, e.g. 10m for most Bluetooth phones. If a server is used, the range is usually unlimited, but the corresponding infrastructure as well as a connection to that infrastructure is necessary.

Table 4.1 provides an overview about social proximity systems, including examples from research as well as commercial applications. The systems are compared regarding their basic technology, identification (is there a global registry of all participants, or are peers identified locally?), peer-sensing capabilities (range and accuracy), whether absolute positioning is used, the required infrastructure, together with the purpose of the system and special features and limitations. Only such proximity systems with a social application (e.g. friend-finders, diaries, or avalanche rescue) are included.

4.2. Proximity Technologies

As we can tell from the tables, there is a large number of different technologies available, that can be used to create social proximity systems. Moreover, they can be combined to compensate for the different limitations of the technologies. Especially, systems intended to work commercially within the urban pervasive infrastructure have to strike a balance between accuracy and availability of the service. The experience gathered with the location research platform Place Lab [151] is in line with several such systems: “Unlike previous location research efforts that focused on maximizing the location estimate’s accuracy, we were willing to trade some accuracy for ubiquity.” ([68], p.32). In the following, the technologies underlying the reviewed social proximity systems are analyzed with their features and limitations.

4.2.1. Radio Frequency (RF) Systems

There are a lot of different RF systems in different scales available for proximity sensing. RF typically passes through walls and other materials. This must be kept in mind, when the distance between people is measured. For many applications, it makes a big difference, if two persons are in the same room or not. When using RF for range es-

System	Technology	Identification	Peer-Sensing	Absolute Positioning	Infrastructure	Purpose	Features & Limitations
Active Badge [191]	IR	Global	No	Room level accuracy	Scanner ($6m$ range)	Coordination of staff in large organizations, e.g. hospitals	None
SPEC [101, 102]	IR	Local	$\leq 5m$ range	No	None	Automatic diary	Generates human-understandable descriptions
Meme Tags [18]	IR	Global	Conversational range	No	IR kiosk	Visualization of relationships and communities	Transmits memes, community mirror
Charmbadge [26]	IR	Global	Conversational range	No	IR kiosk	Conference networking	Flashes on affinity match
Hummingbird [72]	RF	Local	$100m$ range, $\approx 10m$ accuracy	No	None	Interpersonal awareness	None
Avalanche Transceiver [5]	RF	None	$50m$ range, $\approx 1m$ accuracy	No	None	Avalanche rescue	None
Lovegety [81]	RF	None	$5m$ range	No	None	Dating	None
nTAG [143]	RFID	Global	Conversational range	Room level, proximity to exhibits	WLAN access	Conference networking	Real-time audience responses, e.g. quizzes
IntelliBadge [33]	Active RFID	Global	No	Room level accuracy	Scanner ($100m$ range)	Conference networking	None

Table 4.1.: Social proximity systems, part 1

System	Technology	Identification	Peer-Sensing	Absolute Positioning	Infrastructure	Purpose	Features & Limitations
Proactive Displays [120]	Passive RFID	Global	No	In proximity to scanner	Scanner ($2m$ range)	Conference networking	Limited to areas around the displays/readers
Reno [179]	GSM	Local	No	Local symbolic, $100m - 1km$	SMS	Research on privacy in location disclosure	Location is disclosed manually or automatically, sensed automatically
Qiro [155]	GSM	Global	No	$100m - 1km$	Mobile Internet	Show friends' positions on map	None
Just-for-Us [90]	Bluetooth	Global	$10m$ range	In proximity to scanner	Scanner and beacons ($10m$ range), Internet	Encourage new forms of interaction	None
Serendipity [45]	Bluetooth	Global	$10m$ range	No	Mobile Internet	Introductions	None
Nokia Sensor [141]	Bluetooth	Local	$10m$ range	No	None	Socializing, dating	Portfolio on mobile phone
Mobiluck [125]	Bluetooth, GSM	Global	$10m$ range	Place tags, $100m - 1km$	Mobile Internet	Socializing, dating	BT and manual localization
Aka-Aki [2]	Bluetooth	Global	$10m$ range	No	Mobile Internet	Socializing, dating	None
Cityware Facebook [30]	Bluetooth	Global	$10m$ range	In proximity to scanner	Scanner ($10m$ range), Mobile Internet	Understanding of relationships and communities	Bluetooth sensor nodes in Bath
Jabberwocky [149]	Bluetooth	Local	$10m$ range	No	None	Understanding urban social landscape	Visualization of familiar strangers

Table 4.1.: Social proximity systems, part 2

System	Technology	Identification	Peer-Sensing	Absolute Positioning	Infrastructure	Purpose	Features & Limitations
Hocman [51]	WLAN	Local	$100m$ range	No	None	Socializing for bikers	Fast connection and data transfer, when riding the motorcycle
Loopt [109], Mologogo [127]	GPS	Global	No	$10m - 100m$	Mobile Internet	Show friends' positions on map	None
Plazes [152]	IP network address, WLAN	Global	No	Place tags, $100m - 1km$	Internet (wired or WLAN)	Socializing, networking	Additional SMS interface for manual localization
Wearable Face Detector [174]	Optical	None	Conversational range	No	None	Context recognition	Face recognizer may be triggered from input
UbER Badge [57]	IR, Mic, Accelerometer	Global	Conversational range	No	None	Conference networking	Interest detection, affiliation detection
Sociometer [29]	IR, Mic, Accelerometer	Global	Conversational range	No	None	Learning social networks	Influence patterns
Wearable Meeting Annotator [87]	Audio, Accelerometer	Local	Meeting range	No	None	Annotate meeting recordings	Gesture recognition for detection of scenes in meetings
Dodgeball [41], Jaiku [83]	None	Global	No	Place tags, manual localization	SMS	Socializing, dating	SMS interface

Table 4.1.: Social proximity systems, part 3

timation, it is a popular approach to measure the received signal strength (RSSI). The common problem of this technique is, that the signal becomes distorted by people and objects in the environment, resulting in low accuracy in dynamic settings. Another more complicated technique is time-of-flight measurement. The problem of RSSI is avoided, but instead create multipath effects that decrease accuracy. Additionally, the clocks of the devices have to be synchronized.

Cell-ID The GSM infrastructure can be used to provide rough location information, additional to facilitate network access. Chen et al. have shown, that positioning accuracy with a median error of 94 meters can be achieved in downtown Seattle. In the residential part of the city, the median error measured was 196 meters, which can be related to the lower density of GSM towers in that area [28]. The obvious advantages of this technology are, that it is readily available in most populated places around the world and that it can be used with mobile phones. In the US, the E911 directive is implemented by several providers with this technology [50]. While a mobile phone is located by the infrastructure with E911, commercial services like Qiro [155] and Google Mobile Maps use a special program on the phone to estimate the location [58]. Generally, a database with GSM cell tower identifiers and their geographical location is necessary to determine the location in geographical coordinates. Otherwise, only a symbolic value is available. More advanced mobile phone technologies, like UMTS, can basically be used in a similar manner.

GPS The global positioning system was initially built and deployed by the US department of defense to provide navigation for their troops in conflict areas, consisting of a minimum of 24 satellites in earth's orbit. Development began in 1972. The system was completed in 1994 and opened for civil use two years later. In 2004, assisted GPS (A-GPS) was developed to provide increased performance, including minimized start-up time, for mobile phones. The UdSSR built a similar system during the cold war (GLONASS). The Galileo system was initiated in 1999 by the European Union to assure independence from the US. Completion of Galileo is scheduled for 2013. China is building a similar system, called Compass. These systems provide world-wide coverage. Accuracy is heavily dependent on the environment. Reception is usually not possible indoors. Urban canyons decrease accuracy tremendously. A field study of Modsching et al. in the city of Görling, Germany, showed a median error of 15.22m [126]. Another study conducted in Vancouver downtown revealed partially large errors of several hundred meters due to multipath effects, with a median error of 18.4m [99].

Besides its utility for car navigation systems, GPS powers commercial social proximity systems, e.g. Loopt [109] and Mologogo [127]. Olofsson et al. conducted an ethnographic field study at a rock festival to guide design decisions for corresponding systems based on GPS [144]. Like GSM, GPS-based proximity systems are in need of a server infrastructure to calculate proximity. Unlike GSM, the receivers are able to calculate geographic coordinates directly from the satellite signals.

WLAN The IEEE 802.11 series of wireless LAN technologies are popular for laptops and PDAs, and are also being built into some mobile phones. The range of WLAN is typically about 100 meters (outdoors). There are basically three different ways of how a proximity system can be implemented on WLAN.

Ad-hoc In ad-hoc mode, two WLAN interfaces can directly connect, detect their MAC addresses and exchange information, e.g. [51].

Infrastructure local In infrastructure mode, WLAN clients connect to fixed access points. By measuring signal strength and other parameters on the signal, a client can calculate its position based on a database of the geographic information of the access points. There are *deterministic algorithms*, e.g. [6], and *probabilistic algorithms*, e.g. [206], for location estimation.

Infrastructure network The network of access points can also do the location estimation, comparable to the local mode above [171].

The ubiquity of WLAN access points makes them an interesting source for location information. E.g., the Place Lab project [151] contains routines to record the locations of access points and to do location estimation based on previously recorded data. Because WLAN infrastructure is especially high in office buildings, there are commercial systems to take advantage of this situation, e.g. [47]. Location accuracy in such environments is about two to three meters. However, since the signal strength is heavily influenced by people and furniture, accuracy often decreases with changes in the environment. Skyhook Wireless [176] operates one of the largest commercial positioning systems based on WLAN, which is one component of the hybrid location approach in Apple's iPhone 3G, along with Cell-ID and GPS.

Nevertheless, social proximity detection by WLAN is limited, especially for peer-sensing. People use their laptops and PDAs only at a very limited number of different places [121, 171]. When the devices are switched off, WLAN proximity detection does not work at all. WLAN usage in mobile phones is relatively low, mainly because energy consumption is high, compared to other technologies like Bluetooth.

Bluetooth Bluetooth is a low power, short range (10m–100m) RF technology. It is present in most mobile phones today. Its ubiquity makes it interesting for social proximity sensing, especially for applications aiming for wide adoption. The characteristics of Bluetooth are discussed in detail in sections 4.3 to 4.6 of this chapter.

RFID Radio frequency identification systems (RFID) are composed of scanners and tags. The tags are generally very cheap and small, which makes RFID applicable to label products on a large scale. Each tag has its own unique identifier and can transmit it to scanners. There are read-only and read/write tags, that have additional memory to store data. Probably, RFID will replace the optical barcodes in the future, currently identifying all kinds of products. Peer-sensing is difficult to implement, because of the limited range of mobile scanners. Thus, an infrastructure of servers and scanners is usually set up. There are basically two different kinds of RFID systems:

Passive RFID Passive tags do not need their own power source to respond to scanners. Thus, they are very small and cheap. Because there is no need to recharge a battery, they are easy to deploy in a lot of settings with low maintenance costs. The disadvantage is, that the scanning range is limited to a few meters even with powerful scanners. Mobile scanners typically achieve only a few centimeters of range. As part of a conference demonstration, McCarthy et al. installed scanners at posters to automatically identify visitors in their vicinity wearing passive RFID tags [120].

Active RFID Active tags include a power source, enabling them to respond to scanners in a wider range than the passive tags. Therefore, they are prohibitively larger and more expensive to label products. The LANDMARC system explores the use of active RFID for location estimation [132]. IntelliBadge uses this technology for social proximity sensing [33].

RF Custom devices with special radio-frequency (RF) technology can be built to provide peer-sensing proximity detection with a wide range of characteristics. The Lovegety consumer device designed for dating, detects other Lovegety devices in only 5m range and exchanges profile information [81]. The Hummingbird devices have a range of 100m and additionally measure the rough distance within the detection range [72]. The SpotON location system is intended for much smaller scenarios [69]. It has a range of only four meters, but is designed for high accuracy. Transceivers for rapid search and rescue of avalanche victims typically operate in 50m to 100m range and with an accuracy of about one meter to safely localize a victim [5].

4.2.2. Optical Systems

An alternative to the wide range of RF based proximity detection systems are optical systems. A basic difference is, that light does not pass through walls, bodies or other objects. Thus, by measuring proximity with optics, it is possible to take account of the social setting. Additionally, optical systems are generally directional, providing sensitivity for the orientation of people among each other. A disadvantage is that optical systems are generally sensitive to changing ambient light situations, especially to sunlight compared to artificial lighting inside buildings.

IR Infrared is incorporated in many social proximity systems. Senders and receivers are cheap and lightweight. It is possible to combine both in a tiny, mobile device [18]. Senders and receivers can also be separated and either device can be built into a building's infrastructure and the complementary part can be worn by people [191]. Infrared senders are directional, but the light is easily reflected by walls, thus interfering with directionality. Moreover, IR devices consume low power. The sensing range is usually limited to several meters. Infrared devices have to be worn on the outside of the clothing and be fixed to a point in the intended direction to work properly. Often, they are integrated into name tags, affording this requirement naturally.

Video With a much higher complexity, video cameras in combination with computer vision algorithms can be turned into social proximity systems. Singletary and Starner demonstrate the principle with a wearable computer and a video camera mounted on a hat [174]. Their system simply recognizes, whether the wearer is facing another person. More complex—and power consuming—algorithms can also look up faces in a database to identify the conversational partner.

4.2.3. Other Systems

Apart from the large variety of systems based on RF and optics, it is also possible to exploit other technologies and ideas.

IP network address The IP addresses of routers on the Internet are quite static and the locations of many are known. Thus, when connected to the Internet, these addresses can be used to look up the location in a database. Accuracy of this technique is about several kilometers. It can be readily implemented on desktop and laptop computers, as well as on PDAs connecting by WLAN. Plazes.com [152] incorporates this approach.

Manual A very different approach from the previous ones is to abandon automatic detection of proximity and location altogether. Instead, manual proximity systems leave it up to the users to disclose their locations. Privacy problems are thus reduced, because users have good control about the disclosed information. Dodgeball incorporates a database to determine geographical positions from common locations' names and can then determine the distance between its users [41]. The locations disclosed on Jaiku, on the other hand, are up to the interpretations of the users [83].

Hybrid systems By combining several technologies, a more robust or more capable system might be achieved. E.g. the UbER badge [57] and the Sociometer [29] are enhanced IR nametags featuring microphones and acceleration sensors. A microphone might be used to measure properties of conversations, e.g. the directionality to learn about the relationships between the observed persons. Digital cameras with built in GPS receivers or modern camera/GPS phones are another example of such hybrid systems.

There are several other technologies that potentially lend themselves to the implementation of proximity system. The "Active Bat" [192] and the "Cricket" system [154] demonstrate how ultrasound can be used to provide location awareness. A social proximity system could be built on top of these by incorporating a server infrastructure similar to other approaches described above. Ultra-wideband (UWB) provides high accuracy and scalability to track objects in industrial contexts, as demonstrated by the Ubisense product [188]. New technologies, like Zigbee, will probably be an interesting complement, but such devices are not in use in large amounts, yet. Similar to Bluetooth, it includes a device discovery procedure potentially useful for peer-sensing [208].

4.3. Measuring Social Proximity with Bluetooth

In the previous section, we presented a variety of technologies, which have already been used in social proximity systems. Now, we will give reasons for our selection of Bluetooth technology as the main means to collect proximity information for our study.

There are several requirements for the intended collection of data for this thesis. Social proximity is to be measured in urban environments, e.g. in the city, the work place and at home. Collection is further to be conducted on a number of conferences in different countries. The proband should be equipped with a special sensing device, but the majority of the discovered people during the experiment can not be prepared in any way. Further, the exact places the proband will be, and the people he will meet, are unknown in advance. People may be discovered in a variety of contexts, e.g. on the street, in the office, in public buildings, walking, sitting, talking, riding the tram or driving a car. The collected data has to include unique identification of people. Data is to be processed anonymously, but it is necessary to match an encounter with previous encounters to recognize people again and to build a history about every encountered person.

As a consequence, technologies are ruled out, if they require the encountered people to carry an additional device (e.g. custom RF device) or marker (e.g. RFID, IR) for identification. Such an approach is often taken in limited conference or office environments. Video data has several difficulties: correct mounting of a camera is difficult, face identification is intense in processing and it is sensitive to changes in lighting. GPS does not work inside buildings and furthermore, there is no possibility to collect GPS coordinates without a special device or software program. WLAN is probably better suited, but its main use in Laptop computers is not suited for meetings on the street. Moreover, most people use WLAN in infrastructure mode, where peer discovery is limited. Mobile operators maintain large databases of their customers' movements through the GSM network, which could be used for our purpose. The drawback of this data clearly is its low precision.

The pervasive Bluetooth technology in contrast provides a number of advantages for our purposes. With the inclusion in a wide range of mobile phones, Bluetooth has become a wide-spread technology, especially in Europe and North America. Besides the wireless transfer of data over short ranges of approximately ten meters, Bluetooth includes a protocol for the discovery of proximate devices—called *device inquiry*. The limited range results in a satisfactory precision for the measurements of social meetings. While single device inquiries are usually used to find peer devices for communication, periodic inquiries can reveal interesting patterns about the environment, including people and places. Conducting a device inquiry in a crowded place usually reveals an impressive number of peer devices, although the motivation of people for having their devices in visible mode is not obvious. However, it is possible to detect a certain amount of peoples' phones without handing a special device to each of them, which makes Bluetooth appealing for experiments involving a large quantity of persons. The assumption for social proximity detection is, that the presence of a mobile phone indicates the presence of its owner and can thus act as a proxy for a person. Mobile phones are very personal objects and are seldom left behind, especially in public places. It is important to note, that a Bluetooth device reveals its MAC address during inquiry, which is a unique identifier for the device. Additionally, a growing number of fixed devices, like desktop computers, network equipment or HiFi devices are Bluetooth-enabled and can be used to identify places.

Bluetooth proximity detection was already used in a number of studies. Most notably, Eagle and Pentland used it to measure the social network of students and staff on a university campus in an extended experiment with one hundred students over the course of nine months [46]. Hui et al. carried out a similar study during a conference

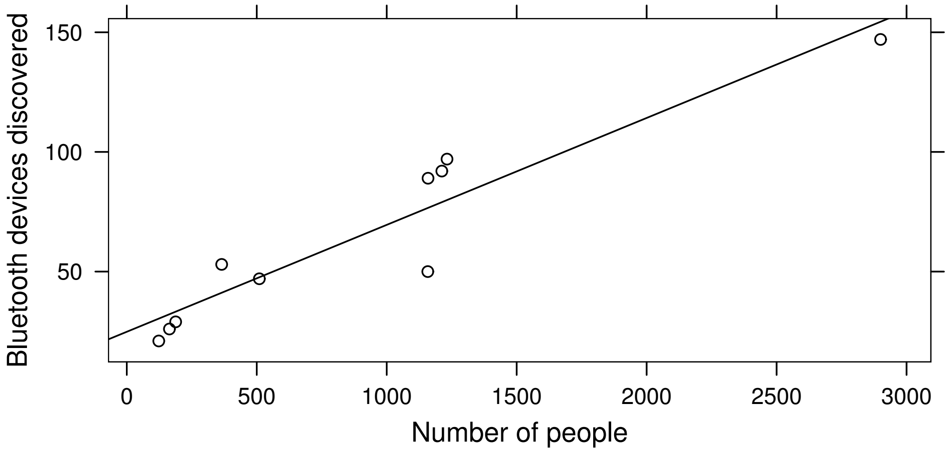


Figure 4.1.: Correlation between number of people and number of discoverable Bluetooth devices across different locations in Bath [7]

with the goal to identify prospects for ad-hoc networking scenarios [77]. Paulos and Goodman on the other hand used proximity detection to measure variables that might indicate the comfort in public urban places [149]. A project at Federation Square in Melbourne, Australia, let users experience and use Bluetooth proximity data to make this otherwise invisible information easy to explore. Among others, they use stationary Bluetooth sensors to detect specific relationships (e.g. acquaintances), as well as the pure amount of people being in places together. Users found this information to be "very cool," "useful" and "fun" ([90], p. 64). Table 4.1 on page 64 also contains a number of commercial examples to take advantage of Bluetooth discovery.

Such information could be used not only to derive social networks, but also to create an automatic diary containing meaningful episodes for the user [103]. Content, like photographs or text messages, could be annotated with information about their context during their time of creation.

However, it is uncertain how the discovered Bluetooth devices exactly relate to people and their behaviors. How many people carry their phones with them? How many do have Bluetooth turned on? Which percentage can be detected? How do these variables change from one place to another? A study by Patel et al. [148] indicates that phones are not even a strong indicator for the exact location of people when they are at home. When away from home (e.g., at the office), the time the phone is kept in arm's length increases to about 70%, and it is plausible that this value is even higher when people are moving through public spaces.

In a study performed in Bath, UK, O'Neill et al. conducted gatecounts to find out how many people in the city had detectable Bluetooth devices with them [145]. In ten sessions at different locations lasting for 30 minutes each, they counted the people passing a conceptual line and scanned for devices at the same time. They used a laptop with four separate Bluetooth interfaces, each inquiring for devices separately. They tested their scanner in a small experiment, conducted before the actual gatecounts. 20 people with detectable devices were sent through a gate to verify that a relatively large number of devices could be detected simultaneously. The results from the ten sessions are shown in figure 4.1. There is a linear correlation between the number of people and the discovered Bluetooth devices, with $R^2 = 0.89$. In total, they conclude that about 7% of people were equipped with discoverable Bluetooth devices in Bath, UK, 2006.

In summary, despite several unknown parameters, Bluetooth has a lot of key advantages compared to other technologies. To assess the data that can be collected with Bluetooth, the remainder of this chapter examines the theoretical parameters and chapters 6 and 7 show empirical results.

4.4. Bluetooth Device Inquiry

A part of the Bluetooth protocol stack is the *device inquiry*. It enables a device to discover other devices in the proximity—usually to establish a connection for data transfer. The discovery process requires active participation of the peer device. It may automatically answer an inquiry request, which can be configured by the user via the Bluetooth visibility option. If it answers, it discloses its device address and device class among others. The address uniquely identifies a Bluetooth device and can be used to recognize a formerly discovered device. The device class distinguishes mobile phones from computers and others and gives vague information about the further capabilities of a device.

The main purpose of Bluetooth technology is to transmit data wirelessly over short distances. Frequency hopping is used to make Bluetooth robust against interference on certain frequencies and to allow multiple separate Bluetooth connections in the same area. Thus, a physical channel is defined by a hopping sequence. Low power consumption is a special design criterion to make it suitable for a wide range of battery powered devices, like laptops, mobile phones and audio headsets. Bluetooth devices are divided into three power classes with a maximum output power of 100 mW (class 1), 2.5 mW (class 2) and 1 mW (class 3) [13]. Transmit distances range from 10 to 100 meters. Nearly all mobile phones are power class 3 devices with a range of approximately 10 meters.

Before data can be transferred over a Bluetooth link, a device has to discover the Bluetooth device address (BD_ADDR) of a peer device. These addresses are 48 bit wide and uniquely identify a device. Since the addresses are assigned by the IEEE

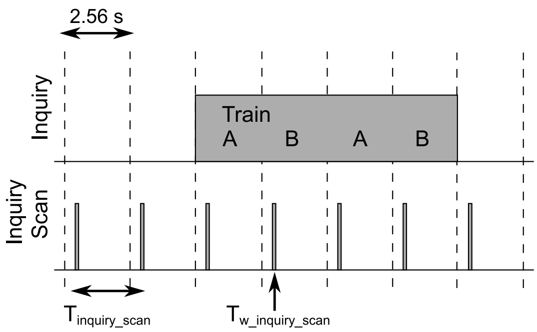


Figure 4.2.: Bluetooth device inquiry timing

Registration Authority and the Bluetooth specification does not define a function to change addresses, devices are uniquely identifiable by their addresses over their whole lifetime.

Figure 4.2 gives an overview over the device inquiry mechanism. To discover another device, a device enters the *inquiry state*. During inquiry, it cycles through 32 frequencies and transmits the inquiry message. The 32 frequencies are divided into two trains, each with 16 frequencies. Each train is repeated for at least $N_{inquiry} = 256$ times, which corresponds to at least 2.56 seconds. The Bluetooth specification recommends to inquire for 10.24 seconds to reliably discover devices in an error-prone environment. The inquiry messages sent do not include any information about the identity of the inquiring device.

In the *inquiry scan state* a device periodically scans for the inquiry messages of an inquiring device. A single scan lasts for $T_{w\_inquiry\_scan} = 11.26$ milliseconds (default value). The time between successive inquiry scans is less or equal to $T_{inquiry\_scan} = 2.56$ seconds. In normal scan mode, each scan is done on a single frequency. Since the 32 inquiry frequencies are split into two trains, a device might scan on a frequency outside the current train. Then, the device is not discovered within the first train of the inquiry, but in the second, when the train is switched. Thus, an inquiry duration of at least 5.12 seconds is necessary to detect a device in an ideal environment. Bluetooth version 1.2 introduced the *interlaced scan* to speed up inquiry. In this mode, the scan is conducted on two different frequencies, ensuring that an inquiry message is received, regardless of the current train of the inquirer. Compared to normal scan, inquiry time can be halved to 2.56 seconds on an ideal medium with the interlaced scan. If inquiry scanning is disabled, a device is invisible to other devices performing an inquiry. However, other Bluetooth functions are still operational.

With the *interlaced scan*, the time it takes a device can be discovered, is strongly influenced by $T_{inquiry\_scan}$. Decreasing this value from its default of 2.56 seconds should also decrease the inquiry time. A modification of the inquiring device is not

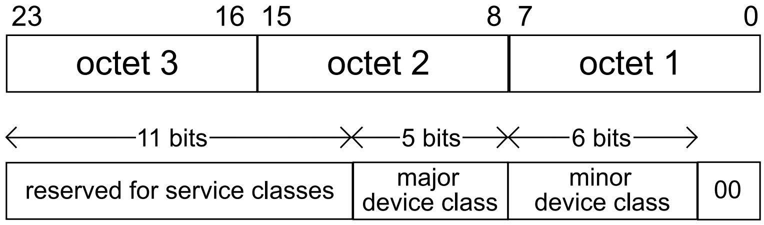


Figure 4.3.: The class of a Bluetooth device is described by three octets (adapted from [12])

necessary. In Bluetooth 1.1 with the *normal scan*, inquiry time is determined by both, the duration of one train of the inquiring device and $T_{inquiry\_scan}$ of the scanning device.

When a device receives an inquiry message in inquiry scan mode, it responds with an inquiry response packet. This packet contains information sufficient to connect to the device, including its BD_ADDR. Additionally, the packet contains information about the class of the device and the services it offers. This is only a rough categorization and may not correspond to the actual services available. To get more information about services, the service discovery protocol can be used in succession. Prevalent device classes are: cellular phone, smart phone,[4] laptop computer and desktop workstation. In contrast to the BD_ADDR, the device class can be changed arbitrarily, provided that the appropriate software is available on a device.

4.5. Bluetooth Mobility Classes

As introduced in the last section, a device transmits its device class during inquiry. An obvious application of this capability is the display of a corresponding icon next to the name of the device. The class is described by three octets in the Bluetooth protocol (see figure 4.3). The descriptor is partitioned into a major and minor class (the service classes are not discussed here). Table 4.2 shows all major and minor classes defined by the Bluetooth specification [12]. It contains a wide variety of different devices, from cellular phones and computers to HiFi devices, wrist watches, jackets and action figures among others. This shows the intention of the Bluetooth SIG (special interest group) to establish Bluetooth as a pervasive short range communication technology.

For our purpose, this classification has interesting implications: Although there is an effort to enable all sorts of electronic devices with Bluetooth, the device classes allow another kind of classification we can exploit to determine, if a specific device can be considered a proxy of a person, a place or an object. We defined five relevant *mobility*

[4]We regard the cellular phone and smart phone to be subclasses of the mobile phone.

Major device class		Minor device class		Mobility
Name	ID	Name	ID	class
Miscellaneous	0		*	U
Computer	256	Uncategorized	0	U
		Desktop workstation	4	S
		Server-class computer	8	S
		Laptop	12	P
		Handheld PC/PDA (clam shell)	16	P
		Palm sized PC/PDA	20	P
		Wearable computer (watch sized)	24	W
Phone	512	Uncategorized	0	U
		Cellular	4	W
		Cordless	8	S
		Smart phone	12	W
		Wired modem or voice gateway	16	S
		Common ISDN access	20	S
LAN/Network Access Point	768	(availability)	*	S
Audio/Video	1024	Wearable headset device	4	P
		Hands-free device	8	U
		Microphone	16	P
		Loudspeaker	20	U
		Headphones	24	P
		Portable Audio	28	P
		Car audio	32	O
		Set-top box	36	S
		HiFi audio device	40	S
		VCR	44	S
		Video camera	48	U
		Camcorder	52	P
		Video monitor	56	S
		Video display and loudspeaker	60	S
		Video conferencing	64	P
		Gaming / toy	72	U

Table 4.2.: Bluetooth device classes and the mobility classification, part 1

Major device class		Minor device class		Mobility
Name	ID	Name	ID	class
Peripheral	1280	Uncategorized	0	U
		Joystick	4	P
		Gamepad	8	P
		Remote control	16	P
		Sensing device	32	U
		Digitizer tablet	36	S
		Card reader	40	P
Imaging	1536	Display	16	S
		Camera	32	U
		Scanner	64	U
		Printer	128	S
Wearable	1792	Wrist Watch	4	W
		Pager	8	W
		Jacket	12	W
		Helmet	16	W
		Glasses	20	W
Toy	2048	Robot	4	P
		Vehicle	8	P
		Doll /action figure	12	P
		Controller	16	P
		Game	20	P
Uncategorized	7936		*	U

Table 4.2.: Bluetooth device classes and the mobility classification, part 2

classes on top of the device classes to be used in the later investigation: wearable (W), portable (P), stationary (S), object (O) and undefined (U) (see table 4.2 for details).

Wearable (W) This mobility class comprises devices, which act as a proxy for their users. The concept of wearable computers was pioneered by Mann [114], Starner [181] and Rhodes [159]. The idea was, that a computer could be worn like clothes (including a head mounted display) and thus be always with the user. The distinguishing concept between wearables and devices like PDAs is, that wearables are *operationally constant*. This means they are never switched off completely, when not in active use, but provide a constant user interface and more importantly, can run processes in the background. Today's mobile phones realize operational constancy in relation to their always-on mode, but not in terms of their user interface. Nonetheless, the implementation of Bluetooth is consistent to the concept of wearables. In this definition, a mobile phone is more wearable than a jacket, because a jacket is regularly changed and not worn throughout each day.

Portable (P) Portable devices, e.g. Laptop computers or PDAs, can be easily carried and are used at various places. In contrast to the wearable devices, they are not operational while being carried (especially in terms of Bluetooth visibility) and not as personal, since they may be used by different people.

Stationary (S) Stationary devices are typically installed in one place and usually not moved regularly. These devices are often wired to the infrastructure or too heavy to carry comfortably. Typical examples are desktop computers, network access points and HiFi devices.

Object (O) Like wearables act as proxies for people, these devices are proxies for mobile objects. A common example is a Bluetooth enabled car audio system, which is fixed to a car. Thus, the car is the mobile object and the car audio system is its proxy.

Undefined (U) For undefined classes, the definition of the Bluetooth device class is not sufficient or ambiguous and can thus not be classified into one of the above mobility classes. This is the case for all the "uncategorized" device classes, but also for devices like a loudspeaker, which could be a small portable device, a large stationary device or built into a car (a mobile object).

With this classification, we intend to cover the prevalent use cases for the different device classes. Because we apply this classification to data collected in the urban environment (see chapters 8 and the following), we have chosen to fit the classification to this context in ambiguous cases (e.g., a jacket is a good proxy for a person outside, but a bad one inside the person's home).

4.6. Distance Measurement with Bluetooth

The device inquiry procedure does not give details about the distance to the peer device, except that it is in communication range (i.e. 10m for most mobile phones). There are no specialized methods to measure the distance within this range in the Bluetooth specification. Still, there is the possibility to use the signal strength as an indicator to distance.

Madhavapeddy and Tse conducted a study to map the propagation of the Bluetooth signal on the basis of the bit error rate (BER) on a connection in an office building [111]. They conclude that Bluetooth is ill-suited for accurate distance measurements. Nevertheless, their results show that it is possible to distinguish rough distances. They measure the BER between two connected devices, thus their method is not directly applicable to anonymous measurements without intervention of the user to establish such a connection.

Patel et al. use a variation of the BER method that is not subject to this limitation [148]. They exploit the service discovery protocol (SDP), which transmits the services (e.g. audio headset services and object push) a device offers. Similar to the inquiry procedure, a Bluetooth device responds to SDP connections without intervention of the user. With increasing BER, the SDP response takes longer, which has similar implications like the method of Madhavapeddy and Tse. The application of Patel et al. does not require accurate distance measurement and is sufficient for their study.

There is also the possibility to make use of the BER during inquiry. The time it takes to detect a device should be in relation with the BER. With increasing BER, inquiry packets get lost and thus inquiry time increases. Another possibility is to count the number of responses which are received from a single device during the period of inquiry. Since devices answer several times to inquiry, the more responses received relates to a lower BER.

Since Bluetooth 1.2, devices report the received signal strength (RSSI) during inquiry. This value also corresponds to the distance between two devices.

4.7. Summary

Our review of social proximity systems has given a wide variety of different technologies, that can be used to detect social activity. However, every technology has its drawbacks and applicability is highly dependent on the application scenario and context of use. The requirement to detect arbitrary people in different urban environments led us to use Bluetooth in our study.

The Bluetooth device inquiry with mobile phones provides features which are well suited for our scenario:

- A relatively large percentage of people carry a Bluetooth enabled and visible mobile phone (about 7% in Bath, UK).
- The low power class of Bluetooth in mobile phones with its small detection radius of approximately 10 meters relates to real-world meetings between persons.
- Bluetooth devices have unique identifiers (the BD_ADDR). With this addresses repeated sightings of the same devices can be recognized.
- Based on the Bluetooth device class, we can determine whether a device is possibly a proxy for a person, a place or an object.

Unfortunately, accurate measurement of the distance between two Bluetooth devices is not supported by the protocol. There are a couple of methods that can be used as approximations. Yet, their applicability and accuracy is questionable. Another disadvantage of the device inquiry is that, regarding to the Bluetooth specification, an inquiry duration of 10.24 seconds is necessary to reliably detect all devices in the vicinity. For fast-moving pedestrians, this duration might be too long for detection.

Part II.

Sensing Proximity with Bluetooth in Urban Environments

5. The WirelessRope Proximity Sensing System

The name of the sensing system used to carry out the experiments for this work has its root in an idea that was developed during a workshop at the Ubicomp '05 conference in Tokyo. While we were exploring the metropolis in a small group of people, we found it hard to be open to the overwhelming amount of impressions, navigate and keep track of the members of the group. The WirelessRope was meant as a technical solution to this problem.

The WirelessRope enables a group to actually feel the boundaries of the group. Like a real rope tying together mountaineers, the WirelessRope gives the urban exploration group immediate feedback (tactile or audio) when a member gets lost or approaches. Thus, everybody can fully engage in the interaction with the environment and cognitive resources for keeping track of the group are freed.

Because of the many limitations of Bluetooth implementations in mobile phones, the initial purpose could not be realized. Nevertheless, the technology proved to be a good starting point for the exploration of social context and guided the design of the graphical user interface (GUI). This chapter outlines architectural design decisions and gives implementation details on the different components of the sensing system. The system comprises three tiers, each with specialized soft- and hardware, so that it can be set up in different environments.

5.1. Sensing System Requirements

The sensing system was designed and implemented with a multitude of goals in mind. The overall purpose was to continually measure proximity between persons with Bluetooth technology. Short range proximity detection was the focus, because it is related to potential contacts between persons. To conduct a study of the intended scale, it was not feasible to equip every possible person with a dedicated device. Furthermore, by selecting the participants beforehand, the results would have been biased. As a consequence, the system had to be able to detect people in general. Of course, no applicable technology could detect every possible person. Instead, a method was required to estimate the percentage of detected people, so that the real amount could be extrapolated.

Most importantly, proximity data had to be collected from an egocentric point of view. A mobile sensing device was thus required that could be carried without being a burden to the user. Complementary, locations had to be included in the study in two different ways. On the one hand side, the mobile sensing device should be able to identify certain places for the correlation of the personal data. On the other hand, stationary sensors were required to monitor people passing by or staying at a place of interest.

The system was further designed to be suited for a long-term study and to be operational in different parts of the world. In particular, the study was conducted at locations in Germany, Italy, Japan and the United States. For the stationary components of the system, easy deployment was a necessity, because it was intended to be used at conferences and exhibitions. The mobile components were also designed for easy deployment, so that it was possible to recruit conference participants to actively participate in the study by turning their mobile phones into sensing devices.

Because there are different possibilities to implement distance measurements with Bluetooth (see section 4.6), the system should also serve as a testbed for the different methods. As a final requirement, the system ought to be equipped with an interactive interface, enabling the spontaneous exploration of the sensed data.

5.2. Related Sensing Architectures

An architecture for a city-wide infrastructure of a "people-centric" sensor network is outlined by Campbell et al. [24]—called MetroSense [122]. Their motivation for an approach, that puts people into the center as carriers of mobile sensor devices, is the increased coverage in comparison with stationary sensors. They find, that a combination of 3,750 mobile sensors and 750 collection points for the data are able to cover an equivalent area of 15,000 static sensor nodes. To facilitate such a large network of sensing devices, they propose a tiered physical architecture. The *sensor tier* consists of mobile sensors attached to persons or objects (e.g. vehicles). These sensors can run small applications and upload their data opportunistically to a device of the *SAP tier* (sensor access point). SAP devices perform similar sensing applications like the mobile tier devices, but are fixed to a location and provide secure and high-availability access to the *server tier*. The server tier is characterized by its high storage capacity and processing power.

Sensing on a city-wide scale was also a method of the Cityware project [31]. Stationary Bluetooth scanners were placed in specific places of the city of Bath to capture pedestrian flow [145]. Compared to the MetroSense architecture, Cityware is limited

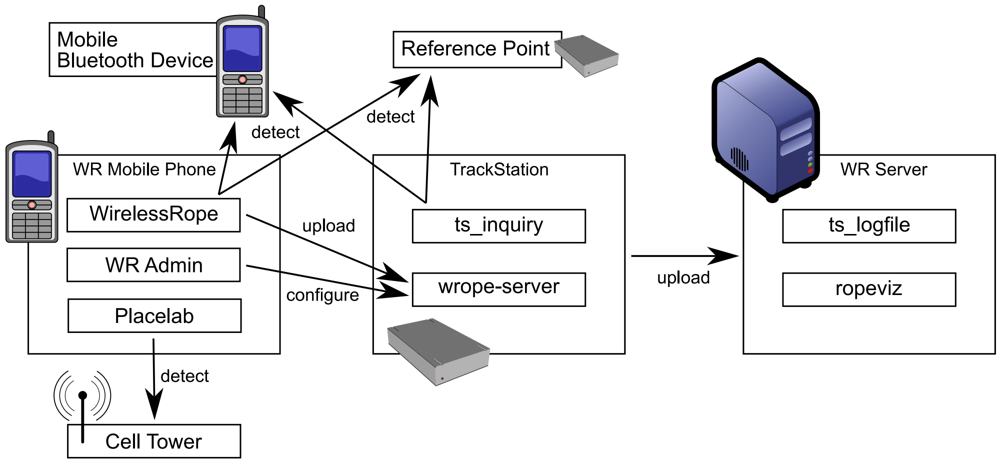


Figure 5.1.: WirelessRope system components

to stationary scanners. However, the Bluetooth enabled phones of pedestrians can be thought of as mobile scanners, because they interact with the stationary devices.

5.3. Architecture of the WirelessRope

Similar to the MetroSense architecture formulated by Campbell et al. [24], the WirelessRope system is divided into three tiers: *mobile*, *stationary*, and *server* (see figure 5.1). Each tier is designed for a different class of devices and a different context of usage.

Mobile tier The primary sensing devices were required to exhibit good mobility. Thus, mobile phones were the obvious choice of hardware. A conglomerate of software components is responsible for the sensing of Bluetooth devices in the vicinity, the detection of the current Cell-ID and the propagation of the sensed information to the second tier via short range radio (Bluetooth).

Stationary tier The second tier is composed of stationary devices, called TrackStations, that can be easily deployed at highly frequented or otherwise meaningful locations, e.g. in conference rooms, train stations or bars. They consist of small, Bluetooth enabled, embedded computers with three primary functions. First, they perform the detection of nearby Bluetooth devices, similar to first tier devices. Second, they act as access points for the mobile devices and receive their sensor logs. The third function comprises the storage and further propagation of the received logs, as well as their own sensor information, to the server tier.

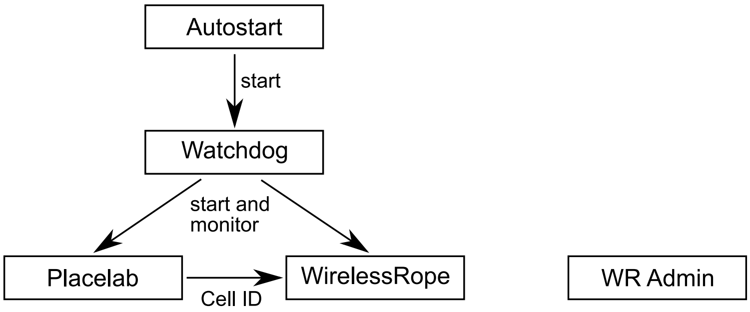


Figure 5.2.: Mobile phone programs of the WirelessRope

Server tier In the WirelessRope system, the server tier comprises exactly one server computer. It processes all sensor information and stores it in a database. A visualization component generates network views from the data, that can be accessed and explored via a web browser.

The largest setup of the whole system was composed of about ten mobile phones, five stationary devices and one server. For small setups, it is also possible to combine the stationary and server tiers into just one device.

Bluetooth device inquiries were implemented on two different hardware and software platforms: mobile phones and PC-like devices. Since the implementations of Bluetooth stacks and devices vary in terms of configurability and capabilities, both were used as measurement devices and could thus be compared in terms of performance.

The source code of the WirelessRope programs is released under the GPL as a contribution to further research. It can be downloaded from the WirelessRope project page at Sourceforge.net [202].

5.4. Mobile Tier

Mobile phones provide the basis for the mobile tier of the WirelessRope. Conceptually, there are two different kinds of devices—*active* and *passive*. The passive ones are phones with its Bluetooth components set to visible mode (see chapter 4), but no additional means for active sensing, storing or processing. The active devices are equipped with special software composed of various components, which are described in the following. Figure 5.2 gives an overview of the components.

On many mobile phones, the J2ME MIDP 2.0 environment is available to run custom programs [84]. With the additional JSR-82 (Java APIs for Bluetooth Wireless Technology [14]) it is possible to access the Bluetooth hardware of a phone. However, there are only limited means to control the inquiry procedure. Most parameters are fixed and cannot be changed. Especially, the inquiry time is fixed to an implementation dependent value. The JSR-82 implementations of the Nokia mobile phones used in this study (Nokia 6680, Nokia 6260 and Nokia 7610) set the inquiry time to the value of eight cycles (10.24 seconds), as recommended by the Bluetooth specification [13]. The main functionality of the WirelessRope mobile tier is based on this technology for portability reasons.

Auxiliary programs are based on the Symbian operating system [182]. Programs for Symbian are written in C++ and allow access to process control and Cell-ID information among others.

5.4.1. WirelessRope J2ME Program

The main program runs on the J2ME platform. It performs periodic Bluetooth device inquiries to collect sightings of surrounding Bluetooth devices. Periodic Bluetooth inquiries are scheduled in random intervals, which is necessary in scenarios with multiple scanning devices. During inquiry, most phones are not detectable by other devices. Thus, random intervals avoid a synchronization of multiple devices, that would prevent the devices to detect each other for an extended period of time. For power saving reasons during long-term use, a random delay between three to five minutes is necessary between successive inquiries. Otherwise, the procedure would drain a phone's battery within a couple of hours. Log data is stored in the RecordStore[5] of the device.

TrackStations are identified by the program based on a fixed address list. When a TrackStation is in Bluetooth range, log data is automatically transmitted. Because this data transfer happens in the background with the user of the device not being aware of it, he might move out of range during this procedure causing an interruption of the Bluetooth connection. Thus, data transfer is divided into chunks that can be transmitted in a couple of seconds to minimize connection problems. Cell-ID information is gathered through the Place Lab component described below. The two programs are connected by a local socket connection.

5.4.2. User Interface

The user interface of the WirelessRope mobile phone program is designed for spontaneous exploration of Bluetooth devices inquiries and implications for one's social context. While common programs which display the Bluetooth neighborhood usually present a snapshot of the current situation, the WirelessRope adds long-term and short-term information about the history of one's personal Bluetooth neighborhood.

[5] The RecordStore is a persistent, key-based storage memory in Java ME.

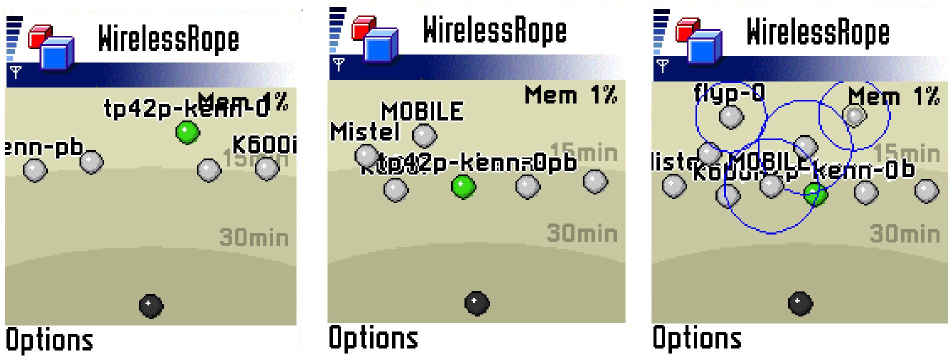


Figure 5.3.: Screenshots of the WirelessRope user interface

Figure 5.3 shows three screenshots of the program. Each discoverable Bluetooth device in the current vicinity is represented by a circle on the screen. The vertical position on the screen conveys the short-term history of each device. The further a circle travels to the bottom of the screen, the longer the device has been in the vicinity. The scale at the right side quantifies the time. During urban walks, users were regularly surprised about circles at the bottom of the screen, since they did not notice the continuous presence of other persons.

For the display of long-term history, the concept of the familiar stranger is incorporated (see section 3.2.3). The WirelessRope counts the number of times a device was encountered, as defined in section 3.4.5. Thus, over time, devices progress on the familiarity dimension:

- from “never met” to “stranger” by first encounter,
- from “stranger” to “familiar stranger” by repeated encounter,
- to “familiar/association” by decision of the user.

Devices are classified into one of three categories and visualized as circles in different colors on the display:

Stranger (gray) All new sightings are classified as strangers.

Familiar Stranger (blue) Strangers which are sighted repeatedly by the proximity sensor are automatically advanced to the familiar stranger category.

Familiar (green) If the user recognizes a familiar person on the display, he can manually add him to the familiar category.

Self (black) The device running the WirelessRope program is visualized on the screen, too.

5.4.3. Place Lab

At Intel Research Seattle, software components for experiments in wireless positioning were developed [100]. One of these components is able to access the Cell-ID information on Symbian OS and provide it to other programs via a local server socket. This component of Place Lab was integrated into the WirelessRope system.

5.4.4. Watchdog

Since the JSR-82 of the phones in the study exhibited stability problems causing application crashes, the Watchdog was implemented to monitor the WirelessRope process. On failures, it reboots the mobile phone. In rare cases, restarts of the WirelessRope without a reboot did not solve the problem, which could be traced back to crashes of the whole Bluetooth subsystem. Watchdog was implemented on Symbian OS using C++.

5.4.5. Autostart

To automatically launch the Watchdog, Place Lab and WirelessRope processes, the Autostart component was implemented (Symbian OS, C++). It closes the loop in case of a failure of the Bluetooth system, resulting in a reboot initiated by the Watchdog. The combination of Autostart and Watchdog proved to be a reliable basis to cope with stability problems of the JSR-82 and Bluetooth in general.

5.4.6. WR Admin

An additional program was developed for the administration of the stationary devices in the second tier, because they may be deployed at places that are difficult to reach and lack a comfortable user interface. The WR Admin J2ME program provides a GUI to enable and disable TrackStations, as well as to query their status.

5.5. Stationary Tier

Comparable to the mobile tier, there are active devices in the stationary tier, called *TrackStations*, and passive devices, called *reference points*.

5.5.1. TrackStations

TrackStations have a similar functionality as the WirelessRope mobile phones and might be installed as additional infrastructure at highly frequented or otherwise meaningful locations, e.g. in conference rooms, train stations or bars. The TrackStations

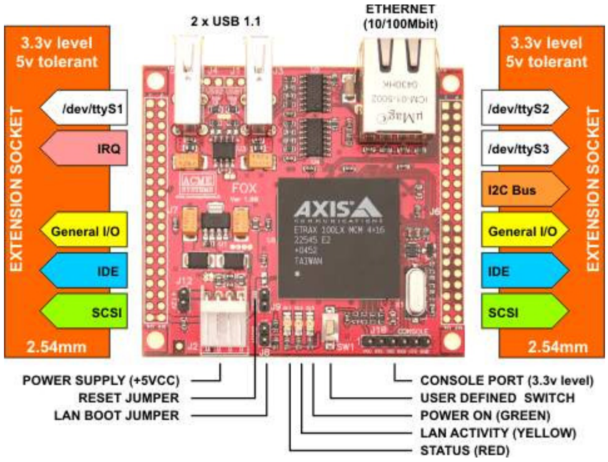


Figure 5.4.: The Foxboard Linux computer is the basis for the TrackStations [1]

automatically record the passing-by of users by Bluetooth device inquiries and can transmit relevant digital tracks to contacts at a later time. By connecting these devices to the Internet, users can also check at which station a contact was seen the last time. By correlating the list of familiar strangers with the list of persons that often visit a station a user may see how much a place is "his kind of place." Paulos and Goodman call this value "turf" [149]. Thus, the TrackStations augment the reach of the WirelessRope at important places. Periodically, these devices collect all log data from the mobile phones and transmit them further to the server tier.

TrackStations consist of small Bluetooth enabled PCs in a box. The Foxboard Linux computers (see figure 5.4) were extended with a module for an SD memory card to store log data and a real-time clock, as well as a USB Bluetooth stick. With the BlueZ Bluetooth stack of the Linux kernel,[6] most of the specified parameters of the inquiry procedure can be modified. The experiments in chapter 6 are based on such modifications.

There are basically two custom software packages installed on the TrackStations (see figure 5.1). The *ts_inquiry* program performs periodic inquiries with several different parameter settings. It builds on the methods given in section 4.6 to implement indicators of proximity within the range of Bluetooth and to quantify proximity:

Inquiry time The time between the start of an inquiry and the reception of the response is recorded.

Number of duplicate inquiry replies During an inquiry, a target device sends multiple inquiry replies. The number of replies received is recorded.

[6] Kernel Version 2.6.12

RSSI Bluetooth version 1.2 introduced the received signal strength measurement accompanying an inquiry.

SDP connect time After inquiry, the SDP protocol is used to retrieve information about the services. The time to connect to the SDP on a target device is recorded. This protocol is useful in this context, because it does not require any user interaction (the user is not prompted to allow this connection). Connection is subject to transmission errors (BER).

SDP browse time When connected via the SDP protocol, the inquiring device can request the services, the device offers. The duration of this request is again subject to the BER and can be expected to increase with distance. However, the amount of data transmitted in response to the request depends on the number of services offered and is thus not comparable between different devices.

The *wrope-server* receives log data from the mobile phones and configures the station, with the *WR Admin* program as its user interface.

If a station is connected to the Internet, log data is uploaded to the server tier automatically. Otherwise, upload can be accomplished manually, by removing the SD card from the station and copying the data with another device.

5.5.2. Time Synchronization

The clocks of mobile phones (and other mobile sensors) are usually not synchronized and often vary a couple of minutes. For data to be consistent, it is important that all timestamps are synchronized. The WirelessRope architecture does not demand synchronized clocks on the mobile tier. The TrackStations of the stationary tier in contrast have to be synchronized to a global clock. This can easily be facilitated, e.g. by NTP (network time protocol).

When a TrackStation receives log data from a mobile device, it performs a simple mechanism to correct the timestamps of the data, if necessary. The mobile device attaches its current time to the sent data. The TrackStation takes the difference between the mobile device's local time and its own synchronized time to correct the timestamps recorded by the local device. Of course, this method always produces a small time difference based on the duration of the transmission. It fails to work, if a mobile device's clock is changed during data collection. Yet, for most cases it works reasonably well within the required time precision.

5.5.3. Reference Points

Reference points do not run the custom software packages of the TrackStations. They are installed in fixed locations and incorporate a Bluetooth module in discoverable

Field	Type	Note
id	int	Identifier for the device
bdaddr	varchar	Bluetooth device address
name	varchar	Bluetooth name
service_classes	int	Bluetooth service classes
major_device_class	int	Bluetooth major device class
minor_device_class	int	Bluetooth minor device class (see section 4.5)

Table 5.1.: Structure of database table “devices”

mode. Thus, these devices have the function to localize the WirelessRope users in space and enable them to recognize formerly visited places.

Most stationary Bluetooth devices can be recognized by their device class (see section 4.5). A place can be identified uniquely by the Bluetooth address of such a device, assumed it is not moved to another place.

5.6. Server Tier

The server tier consists of a single server, which aggregates all data from the mobile and stationary tiers (see figure 5.1). The server is based on Linux with the typical setup of the MySQL database and the Apache web server. We implemented two programs—*ts_logfile* and *ropeviz*—to process and analyze Bluetooth proximity data.

5.6.1. Database Structure

The program *ts_logfile* processes the separate log data files and stores them in a database. The resulting raw dataset is composed of two tables. The table “devices” contains information about all discovered devices, including its Bluetooth device class (see table 5.1). The “log” table contains one entry for every time a device was discovered, including information about the inquiry (see table 5.2).

5.6.2. Network Visualization and Transformation

The combined information can be visualized in real-time as a social network (see section 3.4.1) on a website by the *ropeviz* Java applet. Figure 5.5 shows the output of the program based on data collected on a conference. Nodes are Bluetooth devices, connections between two nodes indicate, that both devices have been in proximity (in contact). The image shows a radial view with one device in the center. Direct contacts are arranged in the first circle, indirect contacts in the outer circle. Users can explore

Field	Type	Note
id	int	Identifier for the log entry
station	int	Identifier of TrackStation, which received this data entry
inquirer	int	Identifier of inquiring device
time	timestamp	Corrected time of scan data
original_time	timestamp	Original (local) time of scan data
sighting	int	Identifier of sighted Bluetooth device
cellid	int	Connected Cell-ID during sighting
areaid	int	Area-ID of the Cell-ID
discoverytime	int	Inquiry time in milliseconds
dup	int	Number of duplicate inquiry replies
rssi	int	Received signal strength (RSSI)
sdpconnect	int	SDP connect time (msec)
sdpbrowse	int	SDP browse time (msec)
sdprecords	int	Number of SDP records transmitted
sdptotaltime	int	Total SDP time (msec)
activity	int	Reserved
dataset	int	Dataset identifier

Table 5.2.: Structure of database table "log"

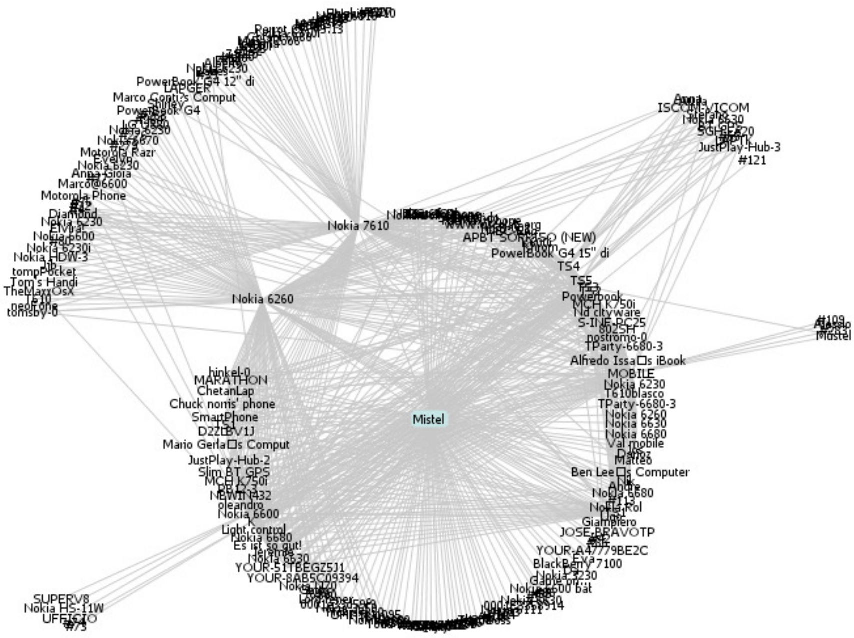


Figure 5.5.: Combined data from mobile and stationary tiers is presented in the form of a social network on a website

their own neighborhood including contacts, regularly met familiar strangers and randomly encountered strangers in this network. It is a tool for personal social network analysis, e.g. to identify common contacts and distinct cliques.

The *ropeviz* program has a number of additional algorithms to transform the contact data for advanced analyses with other programs, e.g. UCINET [16] and Pajek [142]:

proximity This operation calculates 2nd grade encounters from the 1st grade encounters that are directly sensed (see section 3.4.5). Whenever two devices are detected in an interval of less than one minute, a 2nd encounter is defined (demonstrated in chapter 8).

eagle Unique Bluetooth device sightings are counted for each hour of a day. From the result, the eigenbehavior analysis [44] can be run (demonstrated in section 8.6).

dataset This operation creates an affiliation matrix of subsets of data and devices (used in chapter 9).

dayofmonth This is similar to the "dataset" operation, but instead of subsets, an affiliation matrix with separate days is created (also in chapter 9).

mobclass Based on the major and minor Bluetooth device classes, the mobility class of each device is determined according to the mapping of section 4.5 (used in chapters 9 and 10).

cellid This operation creates an affiliation matrix of Cell-IDs and Bluetooth devices (see section 10.4).

sighting A simple matrix of device sightings is created (used in chapter 10).

5.7. Summary

In this chapter, we outlined the various components which constitute the WirelessRope Bluetooth scanning system. Its general architecture is based on a three-tier approach, which separates mobile scanners from stationary access points and the server to accumulate the data. The mobile tier is comprised of components for mobile phones (J2ME and Symbian programs) and features a GUI for spontaneous exploration of social context along its data collection functions. The TrackStations of the stationary tier can perform more sophisticated methods of Bluetooth inquiry along with indicators to quantify proximity. They act as data upload points for the mobile devices. The server tier contains a database for collection of all proximity data and features operations to preprocess and export data for various advanced analysis methods.

6. Experimental Proximity Detection with Bluetooth

The previous chapter introduced the devices and components developed to conduct Bluetooth measurements in various real-world settings. In this chapter, we are going to verify the performance of the system and determine its characteristics as well as its limits in a controlled environment. In the first part, the inquiry procedure is characterized experimentally and the relation between inquiry time and probability of discovery is given. The second part tests the various distance indication methods, which are implemented as part of the *ts_inquiry* program. Repetition in two different environmental conditions shows the limitations of the methods.

6.1. Inquiry Time and Probability of Discovery

To verify the inquiry behavior of the measurement devices, they were tested in a controlled environment. The following experiment shows, that our equipment conforms to the theoretic considerations of the Bluetooth specification, that we reviewed in section 4.4, but also shows slight variations.

6.1.1. Procedure

A TrackStation (see section 5.5.1) was used for this test. A mobile phone with Bluetooth version 1.1 and another with version 1.2 were placed in a distance of one meter. Over 1,000 repetitions of inquiry were performed. For each inquiry, the duration between the start and reception of a response was recorded.

6.1.2. Results

Figure 6.1 shows the durations of device inquiries in the laboratory. As expected, Bluetooth 1.1 exhibits the typical pattern with two peaks in a distance of 2.56 seconds. This is the interval in which frequency trains are changed. Since the *normal scan* is used, only one train at a time is scanned. The median duration for discovery is 2.8 seconds. Bluetooth 1.2 with the *interlaced scan* on the other hand, finishes in one train in most of the times, because both trains are scanned at the same time. Median

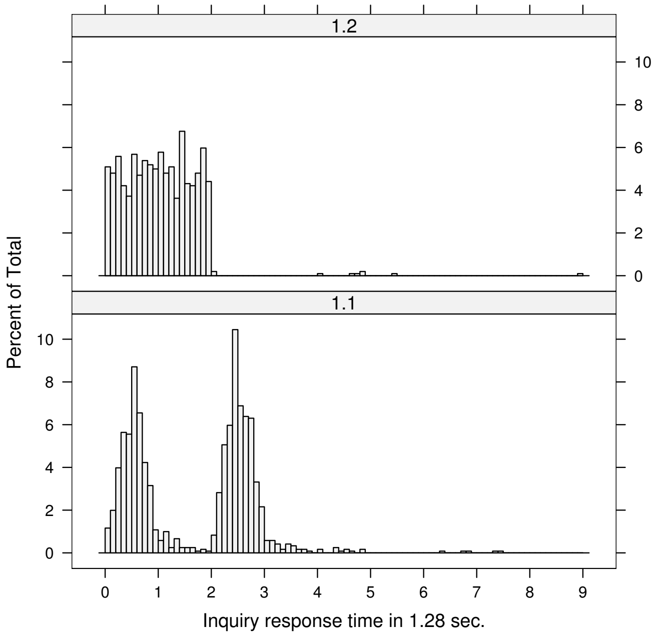


Figure 6.1.: Comparison of discovery times for Bluetooth 1.2 and 1.1 in the lab

duration is thus shortened to 1.3 seconds. The cumulative probability of discovery is summarized in table 6.1. 5.12 seconds were enough to detect a device in at least 92.8% of all cases and 12.8 seconds were needed for 100% probability, which is higher than the recommendation of the Bluetooth specification (10.24 seconds).

6.1.3. Discussion

Not surprisingly, we found two different patterns for the two different Bluetooth versions. This result is problematic for any efforts to use inquiry time for an indication of distance (based on transmission errors with increasing distance), if the Bluetooth version of the inquired device is unknown. The experiment further demonstrates, that even in good conditions (one meter distance, no obstacles between the devices), inquiry fails from time to time. Discovery probability deviates slightly from the recommendation of the Bluetooth specification.

6.2. Impact of the Environment on Distance Measurement

In section 3.2.4, we reviewed different interpretations of the observation that two persons are in proximity. While the Bluetooth range of class 3 devices (about 10m re-

Inquiry time	1.2	1.1
1.28s	49.4	39.7
2.56s	99.1	52.4
3.84s	99.3	84.4
5.12s	99.3	92.8
6.4s	99.8	95.4
7.68s	99.9	96.7
8.96s	99.9	98.0
10.24s	99.9	98.7
11.52s	100.0	99.3
12.8s	100.0	100.0

Table 6.1.: Discovery probability of Bluetooth 1.2 and 1.1 devices in relation to inquiry time

garding to the Bluetooth specification) relates well to the largest range of spatial zones described by Hall [62] (see table 3.2, page 41), the question, if Bluetooth can deliver a finer discrimination between closer zones, is obvious. Is it possible to distinguish intimate distance from personal distance, or social distance from public distance?

Unfortunately, there is no mechanism built into Bluetooth to facilitate the measurement of distance of devices directly. A couple of works have exploited the fact, that transmission errors increase with a weaker signal and thus with increasing distance. In the following experiment we show this effect through various practical methods. We also show, that distance measurement with Bluetooth cannot be used in uncontrolled environments with unknown properties of Bluetooth signal propagation, because the environment has a significant impact on these methods. Thus, such measurements are rendered impossible for our intended use of conducting Bluetooth inquiries in arbitrary locations and environments.

6.2.1. Procedure

Analog to the last experiment, a TrackStation was used to execute inquiries. Five different parameters were measured as indicators of the distance between the inquiring device and the other one (see section 5.5.1 for details on these methods):

- inquiry time,
- number of duplicate inquiry replies,
- RSSI,
- SDP connect time and

- SDP browse time.

To investigate the influence of the environment on the Bluetooth discovery process and the implemented methods for distance measurement, the experiment was conducted separately in two different environments.

Open field There are no metal objects or other obstacles on an open field. The Bluetooth devices were placed on the bare soil.

Corridor Buildings typically have long and tight corridors, with doors on either side.

Bluetooth enabled mobile phones were placed in increasing distances from the inquiring device. The pretest of this setup in the corridor exhibited a significantly extended inquiry range of more than 40 meters, although the Bluetooth class 3 devices have a specified range of only 10 meters. Unfortunately, this exceeded the range of the available corridor but also highlighted the impact of the environment. To clearly measure the differences between corridor and open field, we chose to use Bluetooth modules with removed antennas to limit their range of reception.

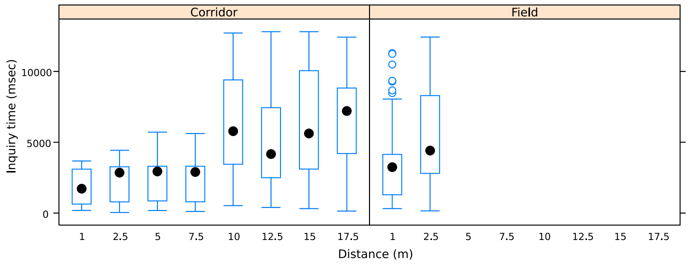
A number of these devices were then placed in increasing distances of 2.5 meters from the TrackStation. 100 repetitions of inquiry were performed, each with every method of the *ts_inquiry* program, as mentioned above. An inactive interval of 13–26 seconds was scheduled between the separate inquiries to eliminate effects of synchronization.

6.2.2. Results

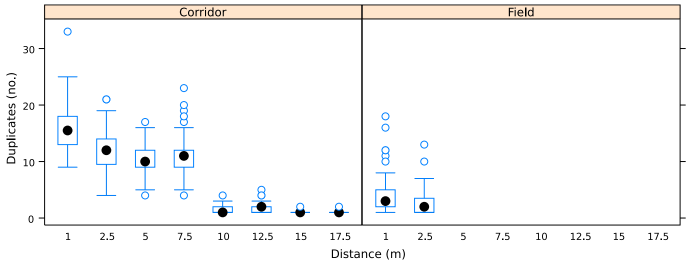
The boxplots in figure 6.2 summarize the results of the experiment. Most of the data validates our assumptions:

- inquiry time increases with distance,
- the number of duplicate inquiry responses decrease with distance,
- signal strength decreases with distance,
- SDP connect time increases with distance, and
- SDP browse time increases with distance.

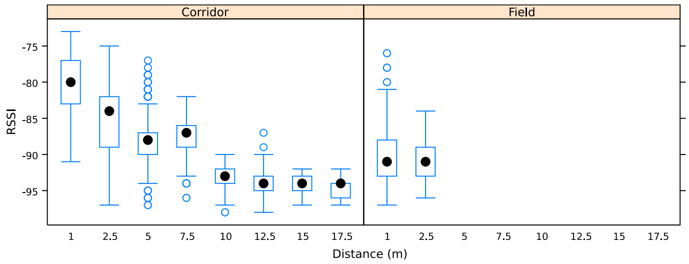
In the corridor, there is an interesting step in the measurements between 7.5 and 10 meters. Thus, it is possible to quickly determine, if a device is more or less than 10 meters away from the scanning device in the corridor. The methods "duplicate inquiry replies" (figure 6.2(b)) and "SDP connect time" (figure 6.2(d)) give the clearest distinctions, regarding to the boxplots.



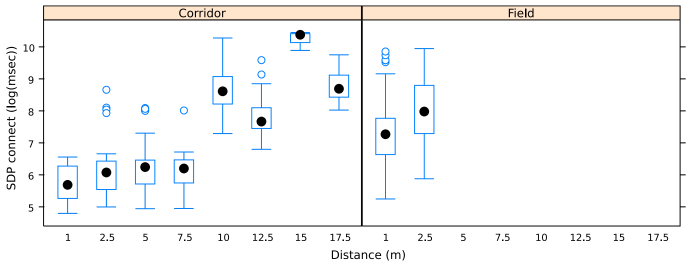
(a) Inquiry time



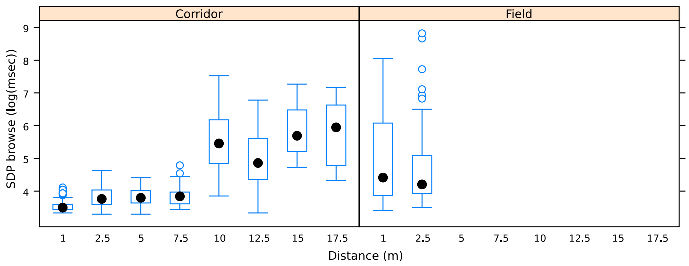
(b) Duplicate inquiry replies



(c) RSSI



(d) SDP connect time



(e) SDP browse time

Figure 6.2.: Bluetooth distance indicator measurements

Unfortunately, the influence of the environment renders these trends insignificant for our study. While the devices in the corridor can be detected in a distance of 17.5m, range is already exceeded at 5m on an open field.

6.2.3. Discussion

The data shows clearly, that the environment has a tremendous effect on the range of Bluetooth inquiries and the distance indication methods. The corridor probably reflects the radio waves and acts like a channel which is able to transmit the signal over a distance that is seven times longer than it is possible on the open field. Unfortunately, we could only quantify this effect with the modified Bluetooth modules (decreased reception), but we can expect similar results for common mobile phones (we were able to detect the class 3 phone in 40m distance in the corridor, limited by the length of it).

The different measurement methods are based on a single effect: the signal strength of the Bluetooth radio system decreases over distance. Thus, with increasing distance, more errors occur, retransmissions become necessary and processes take more time. This explains, that the different distance indicators show a similar behavior.

As a consequence, it seems to be feasible to build a Bluetooth scanner for a given location, that can distinguish between two distance intervals (e.g., close, not so close, not in proximity). Especially, the step in the indicators between 7.5 and 10 meters (corridor) supports this assumption. Thus, the applicability of any of these methods is dependent on the environment and on the devices. If these can be calibrated, as done in [148], different distances can be distinguished with some certainty. However, these methods cannot be used in arbitrary environments, as intended in the study of this work. We can expect the ambiguity in the data to grow, if there are people moving in the places where the measurements are taken. Another drawback is, that all distance indication methods only work on the TrackStations with full access to the BlueZ Bluetooth stack. These methods are not available within the J2ME environment.

6.3. Summary

In this chapter, we have verified the functioning of the WirelessRope measurement devices. Therefore, two experiments were conducted in a controlled environment. In the first experiment, inquiry time and probability of discovery were characterized. Comparison to the theoretic considerations of the Bluetooth specification showed, that the WirelessRope behaves compliantly.

The second experiment examined the effectiveness of a number of distance indication methods within two different environmental conditions (open field and tight corridor). Although the distance indicators respond consistently to different distances, the effect of the different environments render all indicators unusable for unknown environments. We think, that a scanner in a controlled environment could be calibrated to distinguish between a few different distances, but for the intended study with changing environments, these methods are useless. Thus, we will not try to quantify the distance of detected Bluetooth devices in our study and leave this interesting topic for future work.

7. Measurement of Bluetooth Penetration in Urban Places

After testing the technical characteristics of our measurement devices in the lab, we will now turn our attention to the urban environment, that will be part of the final experiment of this work.

People are frequently carrying their Bluetooth enabled mobile phones in their pockets while moving through public spaces. With only a small fraction of these being set to discoverable mode, it is possible to estimate the number of persons in proximity by conducting Bluetooth device inquiries. The proportion might change from situation to situation, with the particular mentalities of the people, cultural differences and the general Bluetooth penetration in a country among others. Some groups of people are more extrovert than others and enable their Bluetooth visibility on purpose. Others are not aware about the consequences and might have it enabled randomly.

Several studies have shown that periodic Bluetooth device inquiry is a rich source of information to implement applications that recognize the (social) activities of the user and to solve technical problems, as in the case of opportunities for ad-hoc networking [44, 45, 46, 94, 145, 25, 77]. The question of how the inquiry results relate to the number of people in proximity is implicit in many of these problems.

In this chapter, we present a method for the measurement of the percentage of people with discoverable devices. We did experiments using this method in Bremen, Germany, and San Francisco, US, suggesting that about 2.3%, respectively 6.2%, are detectable. Further, we show that this measurement can well be implemented on common mobile phones, and that specialized scanners are not necessary.

To count passing people during the Bluetooth measurement, the gatecount method was adopted that O'Neill et al. extended for the Bath study [145]. Therefore, an observer manually counts people crossing a conceptual line through the pedestrian area. The WirelessRope program directly supports this method. It provides a feature to count people by pressing a button on the phone. Compared to counting people on a paper notebook, the exact times of pedestrians passing is recorded by the program in addition.

(a) Lloydpassage in Bremen, Germany

(b) Hauptbahnhof in Bremen, Germany

Figure 7.1.: Gatecount locations

7.1. Survey Locations

Three locations in Bremen, Germany, were chosen that show different characteristics of visitors. Additionally, a location in the US was studied.

Wallecenter Shopping mall distant from the center. There are supermarkets, low budget clothing shops, electronic shops and fast-food restaurants. Visitors are mostly residents. Activities include shopping and spending time in the restaurants. Density of visitors is moderate.

Lloydpassage Shopping mall in the center of the city with high-tech shops, boutiques and fast-food restaurants (figure 7.1(a)). The audience consists of trendy teenagers, old people, tourists and business people. Main activities are shopping and walking slowly along display windows. There is a medium density of visitors.

Hauptbahnhof This is the main station close to the city center with fast-food restaurants (figure 7.1(b)). There are mostly travelers rushing through and teenagers loitering. People are generally in a hurry or waiting for a train. There is a high density of visitors, especially when trains are arriving.

Market Street in San Francisco This is a major street for pedestrians in the city center. There are expensive shops for clothes, coffee shops, fast-food restaurants, music and electronics shops. There are tourists as well as locals. Activities include shopping and sightseeing.

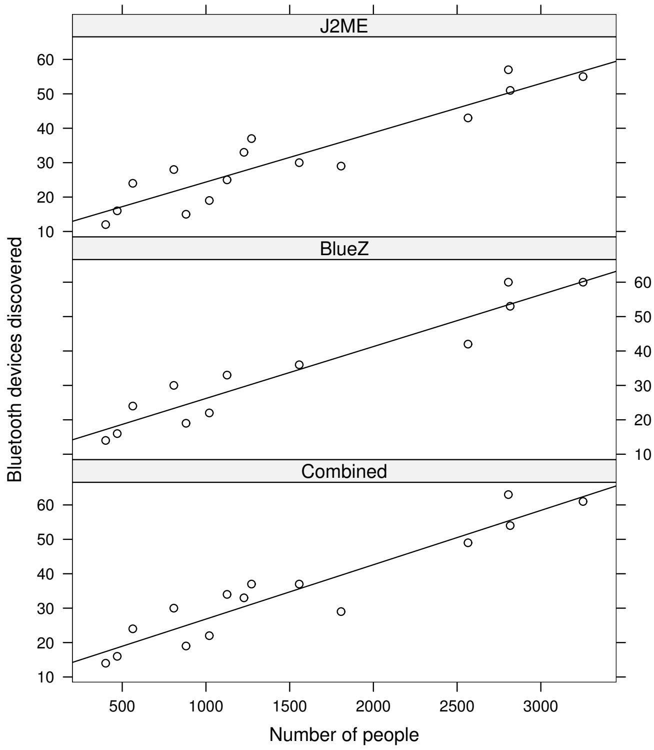


Figure 7.2.: Relation between number of people and number of discoverable Bluetooth phones across different locations in Bremen

7.2. Procedure

Eleven measurement sessions were conducted with both apparatuses simultaneously, TrackStation and WirelessRope on J2ME phone. The four locations were visited for one to two hours on different days during two weeks in August and September 2006 during daytime. An observer was counting all passing pedestrians manually with the mobile phone application.

7.3. Results

7.3.1. Bluetooth in Relation to Pedestrian Count

Figure 7.2 shows the relationship between detected Bluetooth phones and people passing the gates for the three locations in Bremen. The gatecount sessions were divided into 30 minute intervals to make them comparable to the results gathered in Bath

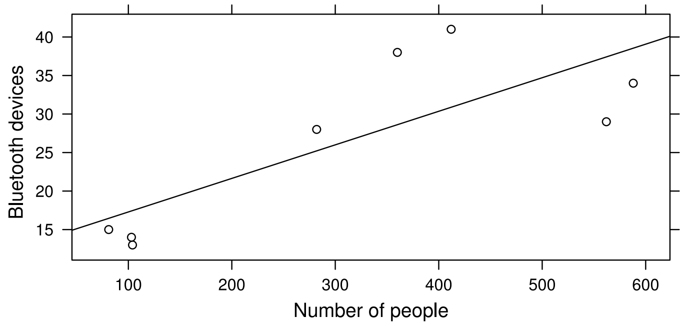


Figure 7.3.: Relation between number of people and number of discoverable Bluetooth phones in San Francisco

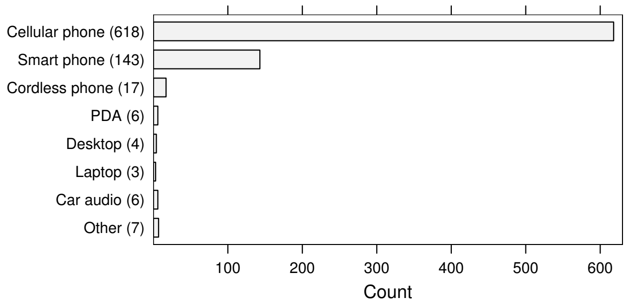


Figure 7.4.: Device classes of discovered Bluetooth devices

[145]. Results were recorded separately for both measurement devices. The third figure shows the combined results of both devices. Three measurements of the TrackStation are missing due to technical problems with the device. Adjusted R^2 of the correlation is 0.85 (J2ME), 0.91 (TrackStation) and 0.88 (combined). Using the same formula for the calculation of overall Bluetooth penetration as O'Neill et al. [145], we get 2.3% for Bremen.

The measurements were conducted with J2ME and BlueZ devices at the same time. When we compare the results of the different devices separately, there is no major difference. Overall Bluetooth penetration calculated from the J2ME data alone gives 2.1%, and the BlueZ data results in 2.2%.

The data from San Francisco shows a higher percentage of Bluetooth devices per person (figure 7.3) with a low correlation of adjusted $R^2 = 0.57$. Here, 15 minute sessions were taken, because less data was collected. Overall, a value of 6.2% of people with discoverable Bluetooth devices was measured.

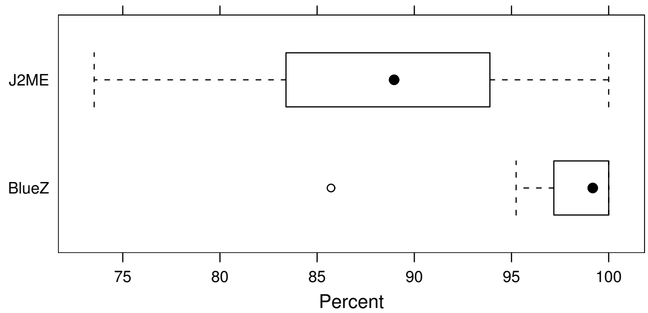


Figure 7.5.: Comparison between the Bluetooth discovery performance of the BlueZ laptop and J2ME phone

7.3.2. Device Classes

According to the device classes acquired during the gatecount sessions, the majority of observed devices were cellular and smart phones (see figure 7.4) with 95%. Cordless phones are tied to a base station and are thus not reliable to detect the locations of people. Since the Bluetooth interfaces of laptops and PDAs are usually deactivated while the devices are carried, these are also not a good measure for passing people. Compared to the amount of phones, these devices only play a minor role. In this analysis, only cellular and smart phones were used. These device classes are good proxies for people, as discussed in section 4.5.

7.3.3. Inquiry Devices

Although there is no large difference between the inquiry devices for the overall calculation of detectable Bluetooth devices per person, figure 7.5 clarifies the difference. When we calculate the percentage of devices discovered by one device, but missed by the other, the BlueZ laptop clearly outperforms the J2ME phone. The BlueZ device discovered 97.7%, while the J2ME device only discovered 88.5% percent of the devices, respectively.

7.3.4. Inquiry Duration

For the gatecount sessions at the selected locations, the short discovery duration of 5.12 seconds was chosen for the TrackStation. Figure 7.6 shows the inquiry times recorded during all sessions, with an unknown number of undiscovered devices. The pattern can be explained by a mixture of both, Bluetooth 1.1 and 1.2 devices (compare to figure 6.1).

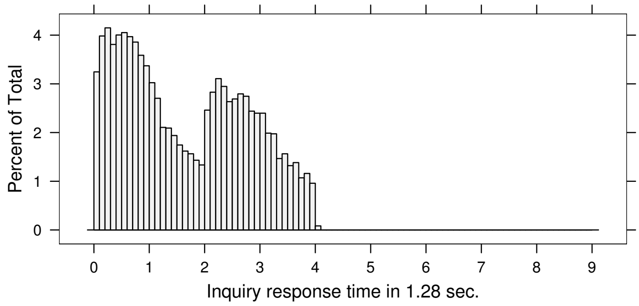


Figure 7.6.: Discovery times with a TrackStation during the gatecount sessions (inquiry time was limited to 5.12 seconds)

7.4. Discussion

Compared to Bath, the percent of persons with discoverable Bluetooth phones is much smaller in Bremen (7% vs. 2.3%). The achieved correlation is comparable. San Franciscans seem to be on a similar level with Bath [145] with about 6.2%. These figures reflect subjective observations about the usage and adoption of Bluetooth technology in the observed cities.

We could not observe a significant improvement in the usage of two scanners as opposed to only one. Although performance of J2ME is lower than BlueZ (2.1% vs. 2.2%), the results indicate that calculation of the amount of people in proximity does not suffer much. The combination of both of our scanners increased the detected part of the people to 2.3%. O'Neill et al. [145] used four Bluetooth dongles at the same time to be sure to discover every device (they tested it with 20 devices simultaneously). The work of Siegemund and Rohs also suggests, that the discovery process degrades with a growing number of devices [173]. However, it is unknown to what extent it can be improved with the usage of multiple scanners.

To further investigate, if and to what extend the inquiry process degrades in the urban environments of our experiments, we made a transformation on our data collected in Bremen. The idea is, that if performance does not degrade with an increasing number of people in the environment, then we should get an equal fraction of people we can detect in high and low frequented areas. Figure 7.7 shows our transformed gatecount data. It plots the number of people in the environment against the fraction of people we could detect by their Bluetooth devices for each of the 30 minute measurements. Although there are a lot of outliers, we can see that indeed the detection rate degrades. When we use this method on the original gatecount data from Bath, the result is even more dramatic (see figure 7.8). With the Bath data, we can observe a steep decrease of detected people, especially between the measurements of 100 to 200

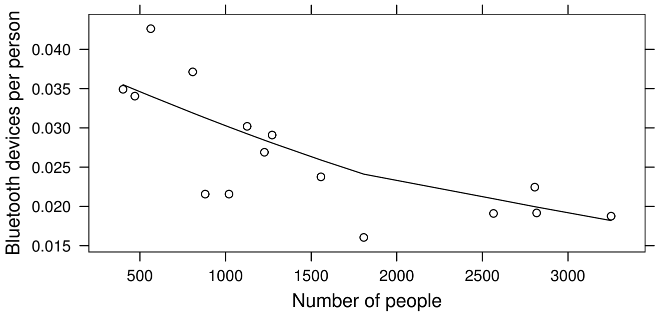


Figure 7.7.: Measured Bluetooth phones per person in relation to the number of people that passed the gates during 30 minute measurement intervals in Bremen

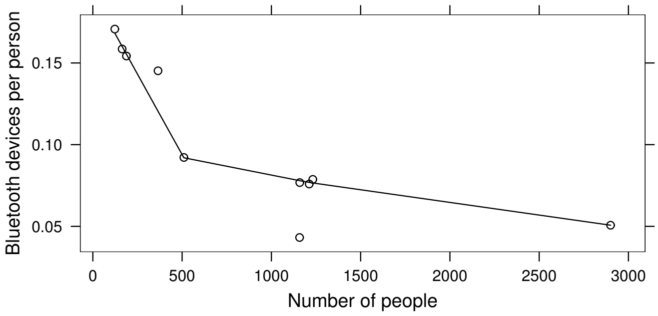


Figure 7.8.: Measured Bluetooth phones per person in relation to the number of people that passed the gates during 30 minute measurement intervals in Bath

people per 30 minute session (about 15% detected) to the measurement of 500 people (about 8% detected). If these observations are not coincidences with different reasons, we can assume that the augmented gatecount method presented by O'Neill et al. [145] underestimates the true amount of people with discoverable Bluetooth devices by a factor of approximately two. Further, we can conclude that the method of using multiple Bluetooth scanners does not remedy the degradation in environments with a lot of people (500 to 1,000 and more during 30 minutes).

When we turn to the different Bluetooth device classes we observed during the gatecounts, cellular phones and smart phones are the prevalent classes in urban environments. This matches well with our intended study, which is mainly concerned with the "wearable" devices we can take as proxies for people (compare section 4.5).

With a wider adoption of Bluetooth 1.2 and later versions in consumer devices, the method described in this thesis can be expected to improve. As shown in the inquiry duration experiment, the device inquiry scan procedure in Bluetooth 1.1 is inferior, resulting in longer inquiry times and less discovered devices. Nevertheless, the presence of 1.1 devices is quite prevalent in our study, as figure 7.6 shows.

7.5. Summary

In this chapter, we applied the augmented gatecount method to measure the percentage of discoverable Bluetooth devices carried by pedestrians in public spaces. With this value, applications can in turn estimate the number of people in proximity by conducting periodic Bluetooth device inquiries.

We conducted a study to measure this value in Bremen, Germany, and in San Francisco, US, and compared it to the results from an earlier study in Bath, UK. Overall, the measured value in Bremen is 2.3%, and 6.2% in San Francisco. We also discovered evidence, that the method might be inaccurate and probably underestimates the amount by a factor of about two.

Part III.

Visualization and Interpretation of Social Context

8. Collection and Properties of the Dataset

In the previous part of this thesis, the development and setup of the technical devices for data collection were discussed in detail. Properties of their function were measured to assess their capabilities. Now, we turn back to the initial idea of this work: how can social context be used as a connecting element between different places and times? How can we visualize social context, if we do not want to limit it to a number of discrete categories?

In this chapter, we introduce a dataset of Bluetooth proximity data. Its collection is outlined and general properties of the dataset are discussed. We demonstrate the construction of an ego proximity network to assess the general structure of the captured dataset and show, how it represents a "map of one's personal social context." We also use the eigenbehavior analysis (see section 3.4.4) on this data, to see how it compares to the network-based methods proposed in this and the following chapters.

8.1. General Properties of the Dataset

The dataset contains Bluetooth proximity data collected in different years and at different geographic locations, which may seem like a random collection at first sight. But since we are interested in the social fabric as the connecting element, it is all centered around a social constant: a single proband, who collected all the data. Thus, the obvious connection between the different times and places covered within the data is the data collector himself. It is important to note, that this social being comes together with his unique social context, which is the subject of our investigation. The proband was thus instructed to act naturally, to do his personal work and leisure activities during the experiment. The author of this thesis took the role of this proband in a self-experiment. During the time of the data collection, he was 30 to 31 years old, research assistant at an institute at the University of Bremen, Germany.

Additional to the data collected by the proband from his personal point of view, selected settings were augmented by a number of stationary scanners set up in locations around the area of the proband. This data allows for a different analysis, showing the social connections between those places and segmenting people visiting those locations into different groups. Another dataset is augmented by location information, collected through the Cell-ID method, enabling us to locate the social contacts of the proband in space. We examine the relation between location and social context in chapter 10.

The whole dataset is partitioned into seven separate subsets, each collected in a different setting. The various settings were chosen to vary widely and exhibit different aspects of social life. They contain data of daily routine, going to work and coming back home, going out on a carnival procession and visiting conferences in different countries. Apart from the obvious differences in geographic location, there are also differences in time: the routine data collection was performed in two successive years. One conference was visited in two successive years, too, and it took place on different continents of the world.

A premise of the data collection was that the identities of the devices that were discovered are unknown. Thus, there is no mapping from device addresses to their respective owners or carriers. Nor do we assume, that the data is complete. The previous sections give an estimation of the expected detection rate.

8.2. The Seven Subsets

The whole dataset is composed of seven subsets, which can be summarized as follows:

#1: Routine05 The first subset was recorded on ten days during June and July 2005 in Bremen, Germany. During the week, the proband was regularly going to work at the university by car. The WirelessRope was configured to record Bluetooth proximity data along with Cell-ID data.

#2: Ubicomp05 The second subset was collected in Tokyo in September 2005 on the Ubicomp05 conference. Additional to the three days of the main conference, the data contains the workshop “Metapolis and Urban Life” (two days) and two other days spent with recreational activities.

#3: Routine06 In February 2006, another routine dataset was collected in a similar manner as the set in 2005. It comprises twelve days, about seven months after the first one, of Bluetooth proximity data.

#4: Carnival06 The day following the Routine06 days, the proband visited a carnival procession in town. Since such an event could give rise to a completely different social situation, it was separated from the other data. It also contains Bluetooth proximity data.

#5: CeBIT06 The CeBIT is a major trade fair of the IT industry and takes place in Hanover, Germany, on a yearly basis. The proband was visiting the fair in his own interests and to visit colleagues from work. Additional to the mobile scanner the proband was carrying, two stationary Bluetooth proximity scanners were installed on two different stands.

#6: PerCom06 On the PerCom 2006 conference in Pisa, Italy, a similar setup as on the CeBIT was realized. A total of five stationary Bluetooth scanners was set up in the venue during the three days of the conference in March. The proband was again carrying a Bluetooth proximity scanner. Several colleagues of him from work were also on the same conference.

#7: Ubicomp06 The last subset of data was collected at the Ubicomp 06 in Newport Beach, Orange County, US. It took place on three days in September 2006, with the proband and his Bluetooth scanner.

8.3. Dataset Collection

The proband was instructed to carry a Nokia 6630 mobile phone with the Wireless-Rope programs installed, as described in section 5.4. The scanning interval was set to approximately 30 seconds (with variation to account for multiple scanners running at the same time). During the experiment, the proband was behaving naturally, doing his own business.

At the CeBIT06 and PerCom06 settings, additional stationary scanners were deployed as described in section 5.5.1. Two scanners at the CeBIT06 and five at the PerCom06 recorded Bluetooth proximity information in the surroundings of the proband. These devices were configured to scan approximately every five minutes, so that there was enough opportunity for the mobile devices to detect them and upload their data in between. Cell-ID information was additionally collected during the Routine05 dataset. It was later correlated to geographic locations (see section 10.4).

After collection of the data, the discovered devices were classified by their mobility classes: wearable (W), portable (P), stationary (S), object (O) and undefined (U). See section 4.5 for the description of this classification.

8.4. Dataset Overview

During the 37 days of the experiment, 1,912 distinct Bluetooth devices were discovered. There are a total of 122,088 Bluetooth sightings. Table 8.1 gives a by day overview about the collected data. “Sightings” denotes the total number of Bluetooth sightings, including duplicate devices. The distinct devices are split up by mobility class (compare section 4.5).

It is obvious that the sightings per day differ significantly, from a minimum of two sightings on 2006-02-08 to 17,105 on 2005-09-12 during Ubicomp05. The same is true for distinct devices. It varies from one on 2006-02-08 up to 593 on 2006-03-10 (CeBIT06). The average number of distinct devices is 65.03 per day. Of these, 43.16 are wearable, 10.57 portable and 3.73 stationary devices. The rest is undefined and no device of the class "object" was found. The total number of portable and stationary devices were usually low compared to the wearables. An exception are the two Ubicomp subsets with a nearly equal portable count compared to the wearables.

8.5. Construction and Properties of the Ego Proximity Network

In contrast to studies like the RealityMining experiment [44], our data is centered around our single proband. In the RealityMining study, a proximity network is constructed on the basis of 1st grade encounters (compare section 3.4.5). This means, that two nodes are connected, when one of them detected the other one. This method would give us quite a trivial network on our data with our proband's node connected to every other one. Such a network would fail to express the richness of the patterns of meetings of our proband.

Instead, we take another approach. Since our proband is present in each interaction (we ignore the stationary devices' data at this point), we can as well exclude him altogether. Instead, we use his observations to connect his contacts with each other, using the concept of 2nd grade encounters. Thus, whenever our proband's scanner detected two or more devices in a time window of one minute, we create mutual connections between all those devices. This transformation is done by the *ropeviz* program (as described in section 5.6.2, by operation "proximity"). For visualization, the Kamada-Kawai energy algorithm [85] is used as implemented by Pajek [142].

The resulting network should be suitable for common network metrics. We verify its properties by measuring its degree of connectedness and apply the small world criteria by Watts and Strogatz [194].

The constructed ego proximity network is shown in figure 8.1. There is a heavily interconnected core in the middle of the graph, although the exact structure is not visible at this level. Around this core, we can see several clusters, laid out like fans. These fan-like clusters are heavily interconnected in themselves, but several such instances seem to be connected to other clusters only through a very few intermediate

				Distinct devices				
Dataset	Date		Sightings	All	U	W	P	S
Routine05	2005-06-29	We	748	25	2	18	3	2
	2005-06-30	Th	351	3	0	1	1	1
	2005-07-04	Mo	2,476	24	2	18	0	4
	2005-07-05	Tu	1,758	27	2	19	2	4
	2005-07-06	We	3,006	24	2	16	3	3
	2005-07-07	Th	3,208	34	2	26	4	2
	2005-07-08	Fr	1,034	67	7	53	1	6
	2005-07-09	Sa	25	8	0	8	0	0
	2005-07-11	Mo	1,844	19	5	12	1	1
	2005-07-12	Tu	3,258	28	5	21	1	1
Ubicomp05	2005-09-10	WS	1,393	64	1	56	4	3
	2005-09-11	WS	5,994	37	0	22	13	2
	2005-09-12	Conf	17,105	115	6	52	49	8
	2005-09-13	Conf	14,356	91	2	40	43	6
	2005-09-14	Conf	10,415	97	4	49	39	5
	2005-09-15	Off	1,292	66	3	15	37	11
	2005-09-16	Off	398	29	3	21	2	3
Routine06	2006-02-04	Sa	85	24	2	21	0	1
	2006-02-05	Su	60	6	0	6	0	0
	2006-02-08	We	2	1	0	1	0	0
	2006-02-09	Th	4,402	81	11	65	3	2
	2006-02-10	Fr	1,451	10	4	3	3	0
	2006-02-11	Sa	1,322	143	5	125	3	10
	2006-02-12	Su	466	8	0	8	0	0
	2006-02-13	Mo	3,773	30	8	19	3	0
	2006-02-14	Tu	8,550	30	7	15	5	3
	2006-02-15	We	4,127	53	9	31	4	9
	2006-02-16	Th	8,190	44	7	31	4	2
	2006-02-17	Fr	9,113	72	6	58	4	4
Carnival06	2006-02-18	Sa	1,532	186	5	174	0	7
CeBIT06	2006-03-10		1,981	593	83	420	73	17
PerCom06	2006-03-14		1,968	61	10	35	11	5
	2006-03-15		2,974	58	9	41	7	1
	2006-03-16		2,096	39	16	13	9	1
Ubicomp06	2006-09-19		491	101	27	35	27	12
	2006-09-20		512	66	17	29	19	1
	2006-09-21		314	42	8	20	13	1

Table 8.1.: By day dataset overview, figures split up by mobility class

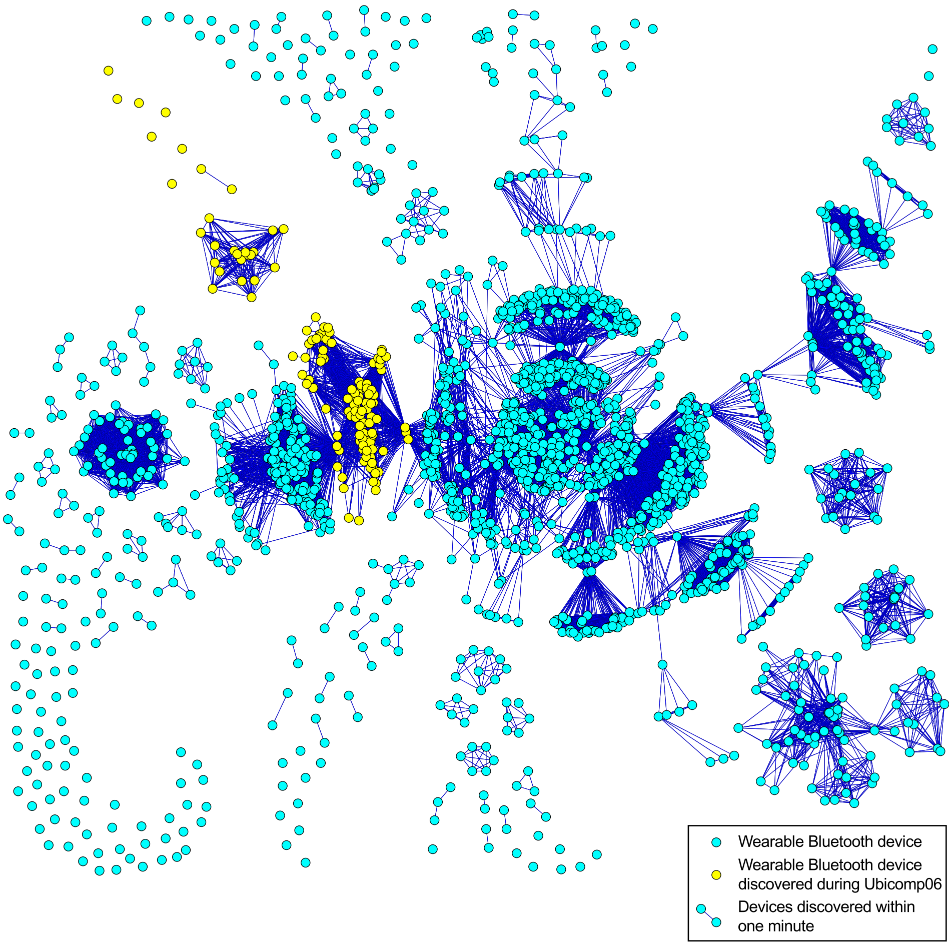


Figure 8.1.: Ego proximity network, nodes discovered during the Ubicomp06 conference are highlighted

nodes. There are also a number of larger clusters, that are not connected to the core and several isolate nodes.

Table 8.2 quantifies the sizes of the components of the network. The core with its connected fans contains 1,384 different nodes, which are about 72.4% of all nodes in the network. The five biggest components contain more than 80% of all nodes. 6.4% of all nodes are isolates.

The small world criterion only applies to connected networks. Thus, we use the largest component of the ego proximity network. Within this component, the geodesic distances are summarized in table 8.3. 84.2% of the nodes are connected by a distance of six or less, 98.3% by a distance of nine or less. The longest path in the component is 15 steps long. The characteristic path lengths and clustering coefficients are presented in table 8.4. They satisfy the small-world criteria $L \geq L_{random}$ and $C \gg C_{random}$ [194].

Size	Frequency	Proportion	Cumulative
1,384	1	0.724	0.724
68	1	0.036	0.760
53	1	0.027	0.787
21	2	0.022	0.809
19	1	0.010	0.819
13	1	0.007	0.826
10	1	0.005	0.831
9	2	0.009	0.840
8	1	0.004	0.845
7	1	0.004	0.848
6	3	0.009	0.858
5	4	0.010	0.868
4	5	0.010	0.879
3	12	0.019	0.897
2	37	0.039	0.936
1	122	0.064	1.000

Table 8.2.: Sizes and frequencies of the components of the ego proximity network

Distance	Frequency	Proportion	Cumulative
1	30,714	0.016	0.016
2	278,924	0.146	0.162
3	371,350	0.194	0.356
4	366,412	0.191	0.547
5	316,524	0.165	0.713
6	247,212	0.129	0.842
7	155,554	0.081	0.923
8	73,414	0.038	0.961
9	41,448	0.022	0.983
10	21,834	0.011	0.994
11	7,762	0.004	0.998
12	2,100	0.001	1.000
13	616	0.000	1.000
14	180	0.000	1.000
15	28	0.000	1.000

Table 8.3.: Frequencies of geodesic distances between nodes in the main component of the ego proximity network

	Proximity network	Random graph
L	4.505	2.676
C	0.830	0.016

Table 8.4.: Characteristic path length L and clustering coefficients C for the proximity network and a random graph with the same number of nodes and average number of edges per node

Eigenbehavior	Reconstruction Accuracy	Cumulative Accuracy
1	0.640	0.640
2	0.114	0.754
3	0.085	0.839
4	0.049	0.888
5	0.045	0.933
6	0.022	0.955

Table 8.5.: Approximation errors for the eigenbehaviors

8.6. Eigenbehavior Analysis

Eagle proposed and applied the eigenbehavior method to the RealityMining dataset to characterize the behavior of his subjects, to characterize group behavior and to determine group affiliation [44]. He interprets the resulting vectors in a similar way as a fingerprint, but related to the behavior of people, instead of the unique pattern of their fingers' skins (see section 3.4.4).

To see, how it responds to our dataset, we conduct the same analysis on our data. The necessary transformation of the dataset is straight forward: we create a matrix with a row for each day in the dataset (37) and a column for each hour of the day (24). Then we sum up the number of distinct devices seen during that time. The *ropeviz* performs this transformation with the operation "eagle." Finally, a principal component analysis is carried out.

The first three eigenbehavior vectors are illustrated in figure 8.2. The x-axis relates to the time of day, color relates to the number of distinct Bluetooth sightings during the respective hour. The first behavior shows increased Bluetooth contacts during the day from 9am to 8pm, with a peak around 12am to 3pm. The second behavior shows increased activity in the morning and the evening. The third one has increased activity during the afternoon and less activity in the morning.

The cumulative reconstruction accuracy of these first three eigenbehavior vectors is 83.9%, the first behavior has an accuracy of 64% (see table 8.5).

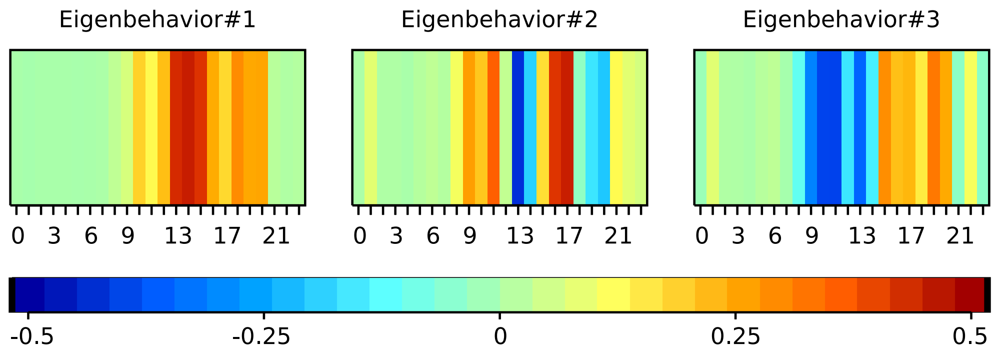


Figure 8.2.: Eagle's eigenbehaviors for the dataset

8.7. Discussion

It is striking, that although the whole dataset was recorded on quite disconnected locations and with two serious gaps in time (more than four months between Ubicomp05 and Routine06, about six months between PerCom06 and Ubicomp06), the proximity network shows a relatively high degree of connectivity (72.4% of nodes). The duration of the separate datasets was short, too, spanning not even two weeks in cohesion.

The fan-like structure of the laid-out network is probably an expression of the different settings our proband was joining. Interestingly, they produced a highly connected pattern. In several cases, only a few—or even a single—devices are responsible for connecting different clusters. The clusters are connected in a hierarchical way, with the core at the top of the hierarchy. There are only a few connections between different branches within this structure.

In retrospect, we can localize our proband on this "map of his social context." When we select any interval in time and highlight the nodes he discovered during that time, it gives us an impression, of where he has been, in terms of social context. Was it in the center or the periphery? Which branch was he on? Which network properties does his current cluster exhibit? Which opportunities might be connected to this position? The highlighted nodes in figure 8.1 illustrate his position during the three days of the Ubicomp06 conference.

Compared to the network-based approach, the eigenbehavior method exhibits different properties of the proband's activities. It is designed to reveal "the repeating structures underlying typical daily human behavior" ([44], p. 89). If we interpret the data of our proband using this method, it results in a rather simplistic view of his behavior (compare figure 8.2): the first vector shows, that our proband was meeting other people between 9am and 8pm, his preferred time for lunch seems to be around 1pm in a public place (the university canteen in the case of the routine subsets). The second behavior can be interpreted as not having lunch in a public place. The third behavior is probably indicative of meeting other people later on the day, but not in the morning.

The number of Bluetooth contacts, which is the basis for the eigenbehavior method, ignores the structure in the data. E.g., whether we are on a normal workday in Bremen or on a conference in Tokyo does not make a clear difference in eigenbehavior, as long as we go to work at approximately the same times and meet people in comparable quantities. The network-based approach, in contrast, highlights the changing structures of human behavior. Even if only a single Bluetooth contact of our proband is examined, we can nevertheless locate him on his ego proximity network, as long as this Bluetooth contact has been discovered before.

The frequencies of the geodesic distances within the network and the measurement of the characteristic path lengths and clustering coefficients give us further measures of the social context. E.g., the higher geodesic distances there are in such a network, the more different contexts an individual connects. The clustering coefficient, in contrast, is indicative of how much the proband is acting in cliques. Moreover, small-world networks have interesting implications on the modeling of infectious diseases. Could this be used to determine one's own risk of becoming infected? Could it tell us not to enter certain social contexts, if we knew, who was infected?

It is important to note, that the resulting network has a lot of holes: there are many sources of error, causing undetected nodes and connections. Chapter 7 explored this issue. On the other hand, there are only a few false positives. There are no devices being detected, which are not there in reality. However, it can be a mistake to assume, that a device stays with the same person all the time. People leave their devices at home [148], they buy new ones, they sell them, lend them or devices get stolen.

Most methods of social network analysis assume complete networks (as far as that is possible), e.g. where single nodes are in the focus of attention. Those methods should not be used on this proximity network, since they are prone to errors in data collection. Methods, taking the whole structure in account, are better suited, because they are generally more robust against missing nodes and links.

Another limiting factor is related to the labeling of the nodes. Since there is no mapping available from Bluetooth devices addresses to their owners, their use is limited.

In practice, the Bluetooth software on the mobile phone caused problems from time to time. Thus, if it was running unattended, data was lost in some cases, e.g. 2006-02-08, where only two sightings were recorded. Reliability could be increased by the introduction of the *Watchdog* and *Autostart* programs (see section 5.4). This approach might be sufficient for the sole purpose of data collection, but if a user wanted to use his phone normally, he would be annoyed by the regular automatic reboots caused by these programs. The high resolution of the Bluetooth scan data of about 30 seconds had the negative effect, that the battery of the phone was usually depleted after half of the day and the proband was required to recharge it during the day or to replace it with a spare battery.

8.8. Summary

In this chapter, we have outlined the dataset for our examination and the procedure for its collection. One proband was using the WirelessRope data collection program in seven different settings, ranging from his daily routine to visiting conferences in different countries in the world. We introduced a method of mapping his social context with an ego proximity network, based on the 2nd grade encounters the proband observed through his scanning device. We compared this method to the eigenbehavior analysis and pointed out the differences between the approaches and their limitations. While the eigenbehavior represents recurring behavior, based on time and the number of people around, the network-based approach enables the localization of scenes within the network and within context. It gives rise to an assessment of the current social context in relation to the whole map.

When we consider our findings so far in relation to the concept of the UPI put forward in chapter 2, we see that a critical mass in discoverable Bluetooth devices is reached. In chapter 7, we measured the number of people which can be detected by Bluetooth device discoveries. In this chapter, we could show that this number is sufficient to create meaningful connections between the people and devices we discovered in the dataset collected by our proband.

9. Social Context of Time

In chapter 2, we identified the temporal structure to be an important aspect of the UPI. Many of the analyses of Bluetooth proximity data we reviewed in sections 3.4.3 and 3.4.4 take the temporal dimension into account, e.g. the visualization of augmented gatecounts (figures 3.17 and 3.18), the discrete Fourier transformation of Bluetooth data (figure 3.21) and the eigenbehavior method (figure 3.22). These analyses establish a relationship between time and Bluetooth proximity data, but they do not regard for the social structure inherent in the Bluetooth proximity data. They have in common, that they are based on the counting of distinct Bluetooth devices during specific periods of time. No relation is constructed to take into account whether the separate Bluetooth devices observed exhibit any relationship among each other or if and when they have been observed before. Thus, the social structure within the Bluetooth data is ignored.

In this chapter, we demonstrate how the understanding of the temporal structure of Bluetooth proximity data can benefit from the social structure—the identities and habits of people—within the data. We introduce a novel method to measure similarity between temporal entities, such as days. This similarity metric can further be used to create discrete clusters of temporal entities. As a result, we obtain sets of temporal entities with similar social structure: in this metric, similarity between temporal entities is high, if the same people are discovered in similar durations; similarity is low, if different people are discovered or the same people are discovered in different durations.

With a slight variation of the method, we are able to substitute social structure with spatial structure. Thus, our similarity metric can also express similarity between temporal entities on the basis of the locations, where the Bluetooth data was recorded, instead of the people that were in proximity. No additional localization technology (e.g. GPS or Cell-ID) is required.

We demonstrate this method on our dataset described in sections 8.1 to 8.3, including Bluetooth contact data collected by the proband with his mobile scanner. Cell-ID information and contact data from stationary TrackStations included in some of the subsets is ignored in this analysis. In two experiments, we apply our method on intervals spanning several days as well as on a day-scale.

9.1. Temporal-Cluster Method

This method operates on Bluetooth proximity data collected over an extended period of time (e.g. several days). It can cope with data collected by both, a mobile scanner or a stationary scanner. Data collected by a mobile scanner can be analyzed for temporal similarity by social structure and by spatial structure. Either of these options is selected by filtering Bluetooth devices for their mobility class. For a fixed scanner, only the analysis by social structure is available. Spatial structure is trivial in this case. First, a two-mode affiliation network is created to put temporal entities in relation with Bluetooth devices. This network is then transformed to a one-mode network representing relations between temporal entities only. We then use the concept of *social positions* (as introduced in section 3.4.1) and hierarchical clustering to create discrete sets of temporal entities. The specific steps of the temporal-cluster method are the following:

1. Bluetooth contact data usually contains a wide range of different Bluetooth devices, including mobile devices (e.g. mobile phones) and stationary devices (e.g. desktop computers). We may either choose to include all devices in the analysis or filter the devices by their mobility class to focus on a specific aspect. To establish a connection between time and social context, we select devices of the wearable mobility class. To establish a connection between time and location, devices of the stationary mobility class are selected (see section 4.5 for a definition of mobility classes).

2. We choose the temporal unit that we want to analyze in the dataset of Bluetooth contact data. E.g., we may select to analyze by day, by week or other scales we might be interested in. From the temporal unit of interest, we create a set of discrete temporal entities, which are present in our dataset (e.g. the days 2006-03-14, 2006-03-15, 2006-03-16).

3. A two-mode affiliation network of Bluetooth devices and the temporal entities chosen in the previous step is constructed. Thus, there are two different kinds of nodes: temporal entity nodes and device nodes. In this network, a device is connected to a temporal entity, if the device was detected during the endurance of the temporal entity. The strength of the tie corresponds to the number of sightings in this period. The affiliation network is represented by a matrix $A = (a_{ij})_{i=1...n,j=1...m}$ with a row for each device (n is the total number of devices) and a column for each temporal entity (m is the total number of temporal entities).

4. In the next step, we create a one-mode network characterizing the connections between the temporal entities, based on the devices the temporal entities have in common. From the affiliation network A, the one-mode network is defined as $D = (d_{ij})_{i=1...m,j=1...m}$, where

$$d_{ij} = \sum_{k=1}^{m} min(a_{ki}, a_{kj}) \tag{9.1}$$

   Thus, the temporal entity nodes get directly linked by the device nodes that are discovered during the temporal entities. For non-valued matrices, a matrix multiplication $D = A^T A$ (where A^T is the transposed matrix) is commonly used. Instead, we use the minimum here, because the relation is valued and contains the number of sightings per device and temporal entity. The minimum gives us the overlap of sightings. Thus, in the resulting network, temporal entities are more related, if the number of sightings between the same devices are similar. With the bigger difference in sighting count between the same devices, or different devices, the relation becomes looser.

5. The diagonal of D contains the total number of sightings per temporal entity. We use these values to normalize the matrix. A minimum method is used for normalization to cope with failures during data collection and also with the different durations of data collection. The normalized matrix $D' = (d'_{ij})_{i=1...m,j=1...m}$ is calculated by

$$d'_{ij} = \frac{d_{ij}}{min(d_{ii}, d_{jj})} \tag{9.2}$$

6. As an intermediary result, we can visualize the network defined by D'. The network contains one node for each temporal entity and connections indicating the overlap between the temporal entities. The strengths of the connections is represented by the width of the connecting lines in the network diagram. Depending on the number of temporal entities and the network structure, the resulting network picture can give us a clear image of the temporal structure within a dataset.

7. We examine the temporal network further by measuring the structural equivalence (see section 3.4.1) between the separate temporal entities. Pearson correlation (equation 3.3) is chosen as a measure of equivalence for this case, because it focuses on similarity in pattern of the ties between the temporal entities. Other measures are used to compare the strength of the ties directly. Thus, Pearson correlation is more tolerant of missing data (compare section 3.4.1). After applying the correlation, we obtain a matrix $S = (s_{ij})_{i=1...m,j=1...m}$ containing the pairwise similarities between the temporal entities.

8. Finally, discrete groups of similar temporal entities can now be created by applying single-link hierarchical clustering on the similarity matrix S. The result is a dendrogram giving a continuous refinement of groups of temporal entities.

The following sections demonstrate this method on different temporal units and with different mobility classes.

9.2. Subset Network Experiment

In this section, we apply our temporal-cluster method to generate a subset network—a highly condensed view of our dataset described in chapter 8. The ego proximity network in figure 8.1 has shown a high degree of connectedness, but the specific connections between the subsets recorded in different settings remain covert. We do not apply the whole method given in the previous section. Because there are only seven subsets, we will interpret the network created in step six of the method directly.

The raw dataset is taken as the basis for the subset network. As for the ego proximity network, only data collected by the proband is used, data from stationary scanners is omitted at this point. For this experiment, we select the subsets of data, as described in section 8.2, comprising a varying number of days each, as the temporal unit of analysis.

From the data, we build a two-mode affiliation network containing subsets and devices. Thus, there are two different kinds of nodes: subset-nodes and device-nodes. In this network, a device is connected to a subset, if the device was detected by the proband as part of the subset. The strength of the tie corresponds to the number of sightings.

The affiliation network is represented by a matrix $A = (a_{ij})_{i=1...n,j=1...m}$ with a row for each device ($n = 1,911$) and a column for each subset ($m = 7$). The corresponding transformation is done by *ropeviz*, with the operation “dataset” (see section 5.6.2).

In the next step, we create a one-mode network characterizing the connections between the subsets, based on the devices the subsets have in common. From the affiliation network A, the one-mode network is a matrix $D = (d_{ij})_{i=1...m,j=1...m}$, where

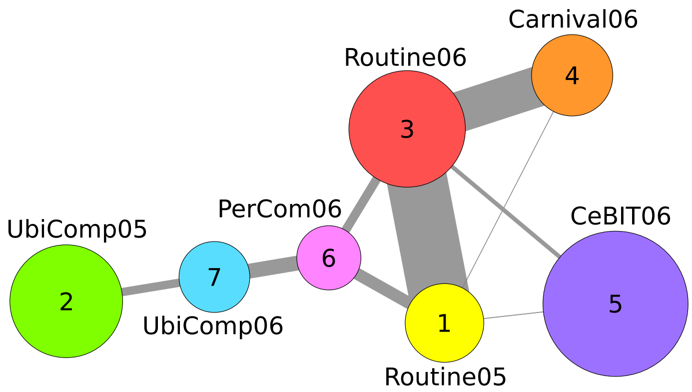


Figure 9.1.: Network showing the relations between the subsets of the dataset

$m = 7$. Thus, the subset-nodes get directly linked by the device-nodes that are discovered in the subsets. Matrix D is then normalized. The normalized matrix D' is:

$$D' = \begin{bmatrix} 1 & 0 & 0.254 & 0.003 & 0.004 & 0.051 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0.035 \\ 0.254 & 0 & 1 & 0.170 & 0.018 & 0.037 & 0 \\ 0.003 & 0 & 0.170 & 1 & 0 & 0 & 0 \\ 0.004 & 0 & 0.018 & 0 & 1 & 0 & 0 \\ 0.051 & 0 & 0.037 & 0 & 0 & 1 & 0.062 \\ 0 & 0.035 & 0 & 0 & 0 & 0.062 & 1 \end{bmatrix}$$

The first row and the first column corresponds to subset #1 (Routine05), the second row and column to subset #2 (Ubicomp05) and so on. The full numbering is given in section 8.2. We visualize this matrix as a network, as shown in figure 9.1. The seven nodes represent the subsets of our data, the number labels of the nodes correspond to the row and columns order of the matrix. The ties represent the overlap in datasets as calculated above. For the purpose of visualization, the size of the nodes corresponds to the number of distinct devices in a subset, as given by vector

$$v = \begin{bmatrix} 174 \\ 358 \\ 382 \\ 186 \\ 593 \\ 116 \\ 141 \end{bmatrix}.$$

The subset network (figure 9.1) shows a very clear overview about the connections between the subsets of data. There is a strong connection between the Routine05 and Routine06 data, despite the difference in time of seven months. The Carnival06 subset is also connected strongly to Routine06.

The network also bridges differences in space: PerCom06 (Pisa, Italy) is well connected to both Routine datasets (Bremen, Germany). Further, there is a connection between PerCom06 and Ubicomp06 (Orange County, US), which is connected to Ubicomp05 (Tokyo, Japan).

9.3. Day Network Experiment

Next, we create a similar network as described in section 9.2, but instead of the rough subset scale, a finer day scale is chosen. Additionally, we use clustering to form concrete groups of days. These groups consist of days, during which our proband was moving in similar contexts—socially or geographically.

Again, we take the raw dataset from the personal scanner of our proband as a basis. A two-mode network consisting of day-nodes and device-nodes is constructed ("dayofmonth" operation of *ropeviz*, section 5.6.2). As a result, a device-node is connected to a day-node, if the device was sighted on that day. The strength of a tie corresponds to the number of sightings during a day.

We use equation 9.1 to transform the two-mode network to a one-mode network and equation 9.2 to normalize it. As a result, we obtain the network shown in figure 9.2, with the sizes of the nodes corresponding to the number of distinct devices discovered per day. The day network expresses the relation between the separate days.

Subsequently, we examine the day network further by measuring the structural equivalence between the separate days. The program UCINET [16] was used to carry out the analysis. Pearson correlation is applied to calculate matrix S containing the pairwise similarities between the days.

Discrete groups of similar days can now be created with single-link hierarchical clustering. The result is the dendrogram shown in figure 9.3. The dendrogram gives us a continuous refinement of groups. Starting at the right side, all days are in one single group. By tracing the tree-diagram to the left, each split creates an additional group, beginning with the most meaningful split and ending in the most trivial one with only slight differences between days. Since our dataset is composed of seven subsets with quite distinct settings, we choose this number of groups to determine the split. The blue vertical line in the dendrogram represents this split and the days are thus separated into the groups indicated by the dashed blue horizontal lines. Our choice of seven groups results in a level of $a = 0.3$. This level is calculated by UCINET during

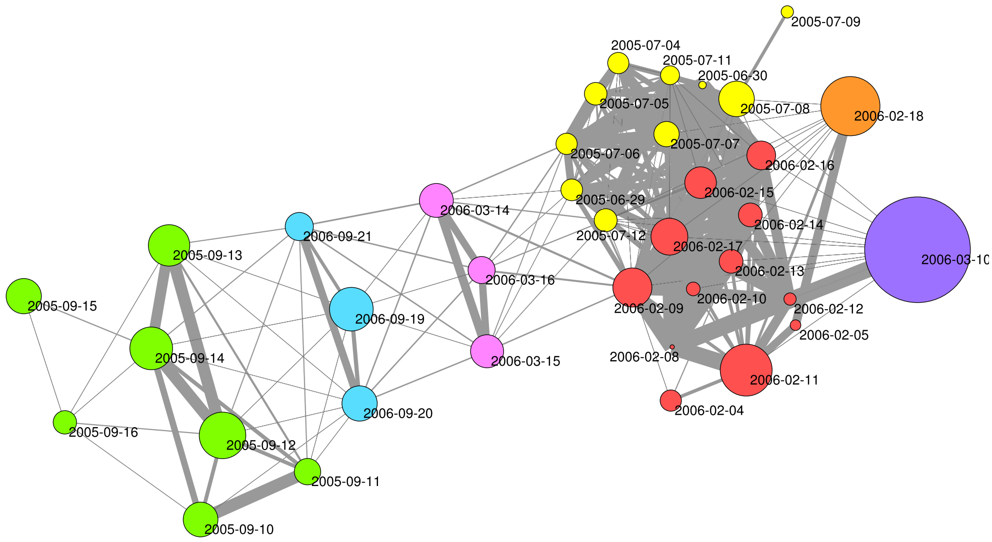


Figure 9.2.: Day network (colors of the days correspond to the subsets, compare to figure 9.1)

the clustering. For single-link clustering with a similarity relation, a level of $a = 0.3$ means that every item in a cluster is at least 0.3 units similar to at least one other item in the cluster.

As a final step, we compare how this result relates to the different mobility classes of devices defined in section 4.5. Thus, we repeat the whole procedure, with the raw data filtered for wearable devices, stationary and portable devices (filtered by *ropeviz*, operation "daymobclass," section 5.6.2).

Partitioning of the resulting dendrograms is done with the same level of $a = 0.3$ as in the first one to make them comparable. As a result, we receive the groups of only the wearable mobility class (figure 9.4), the stationary class (figure 9.5) and the stationary and portable classes combined (figure 9.6). Table 9.1 summarizes the groups for different mobility classes. After applying the mobility class filters, some of the days become disconnected from the other days, because there is no overlap in the discovered devices. The respective days are shown, but grayed out in the dendrograms. In table 9.1, the corresponding cells are left empty.

The day network shown in figure 9.2 shows the same general structure as the subset network in figure 9.1. The subsets of data are split into the separate days they are composed of and the internal structure of the subsets becomes visible as well as their connection pattern to days of other subsets.

When we examine the green Ubicomp05 day-nodes of the network, we can observe a clear picture of those days: the 10th and 11th are well connected (workshop days). Then the three days of the main conference are strongly connected, but a little weaker

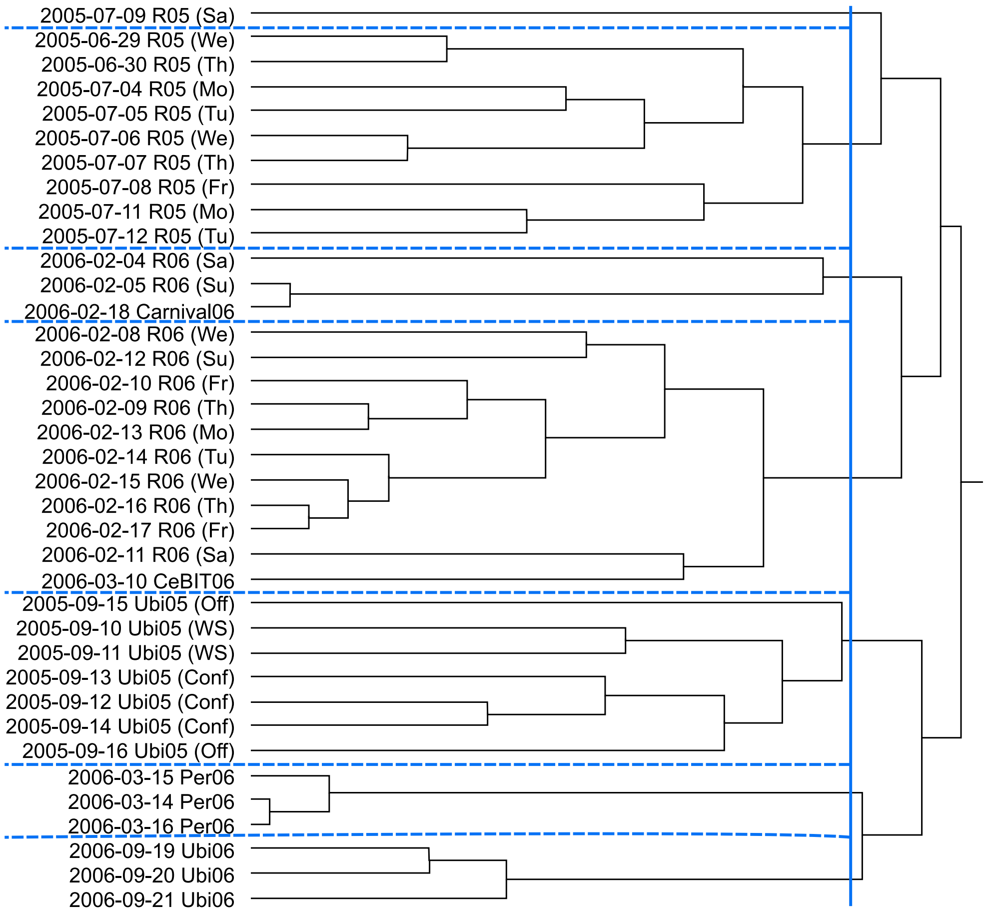


Figure 9.3.: Day clusters, all mobility classes

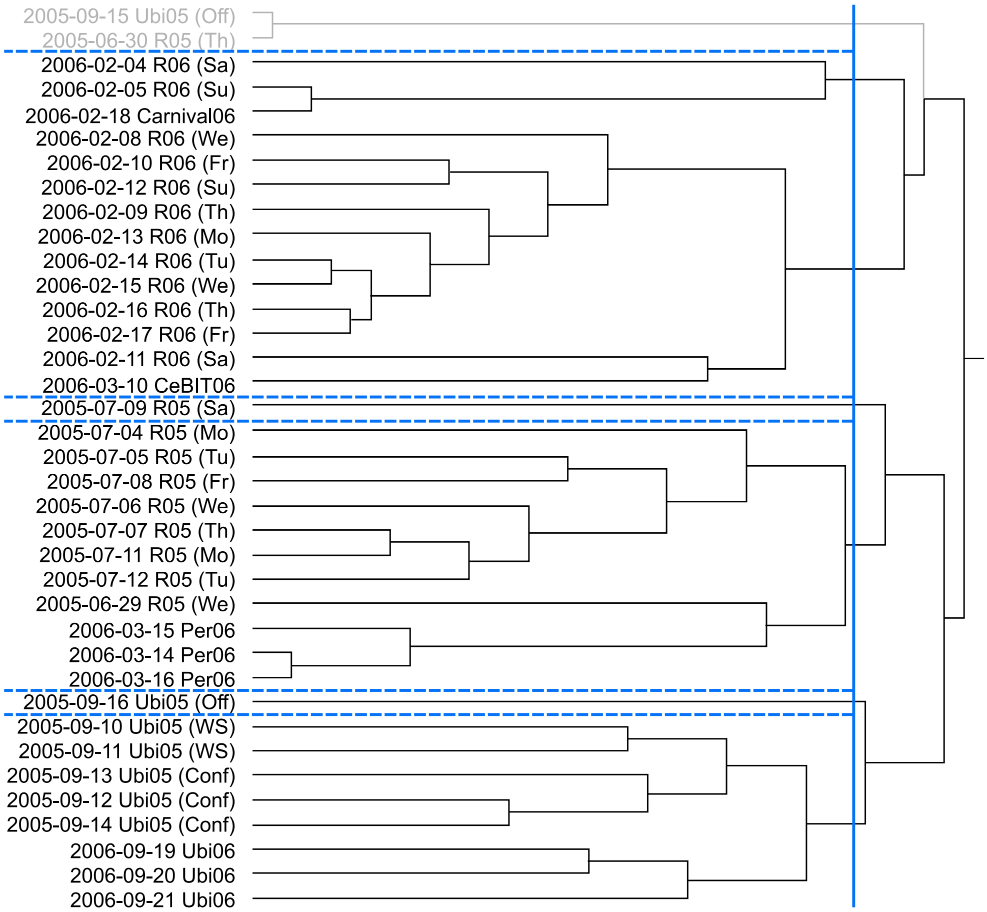


Figure 9.4.: Day clusters, wearable mobility classes

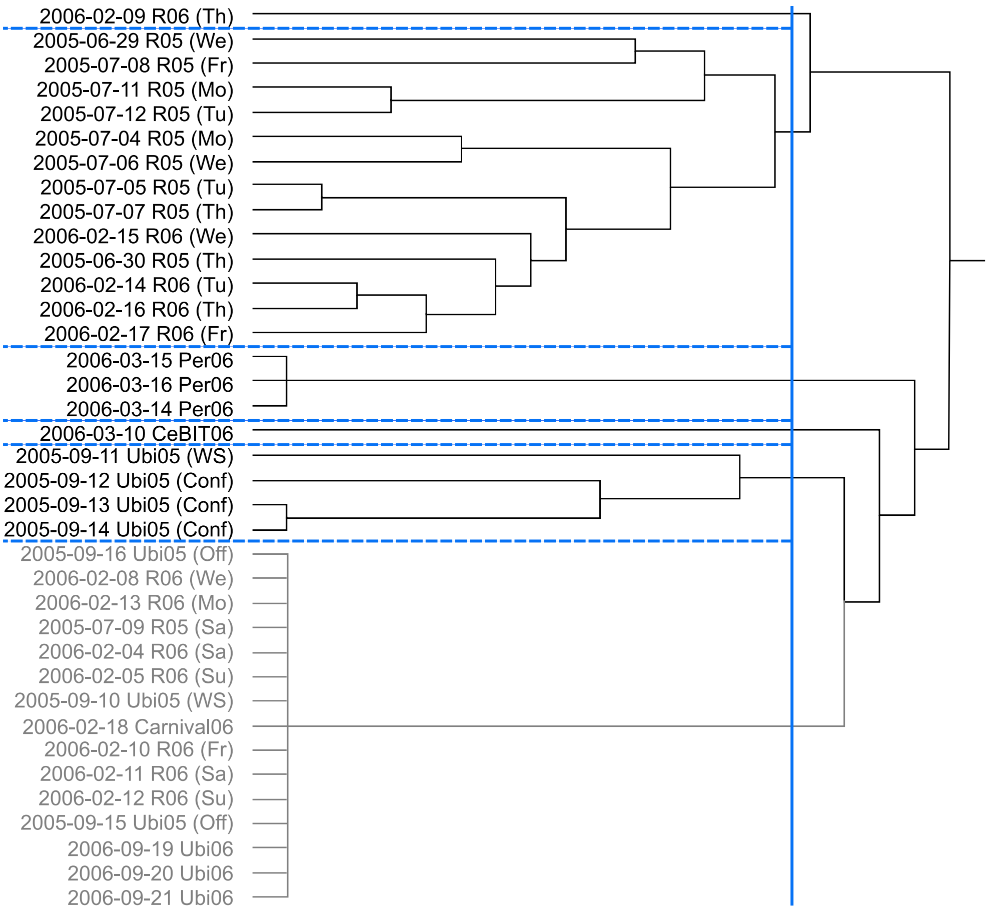


Figure 9.5.: Day clusters, stationary mobility classes

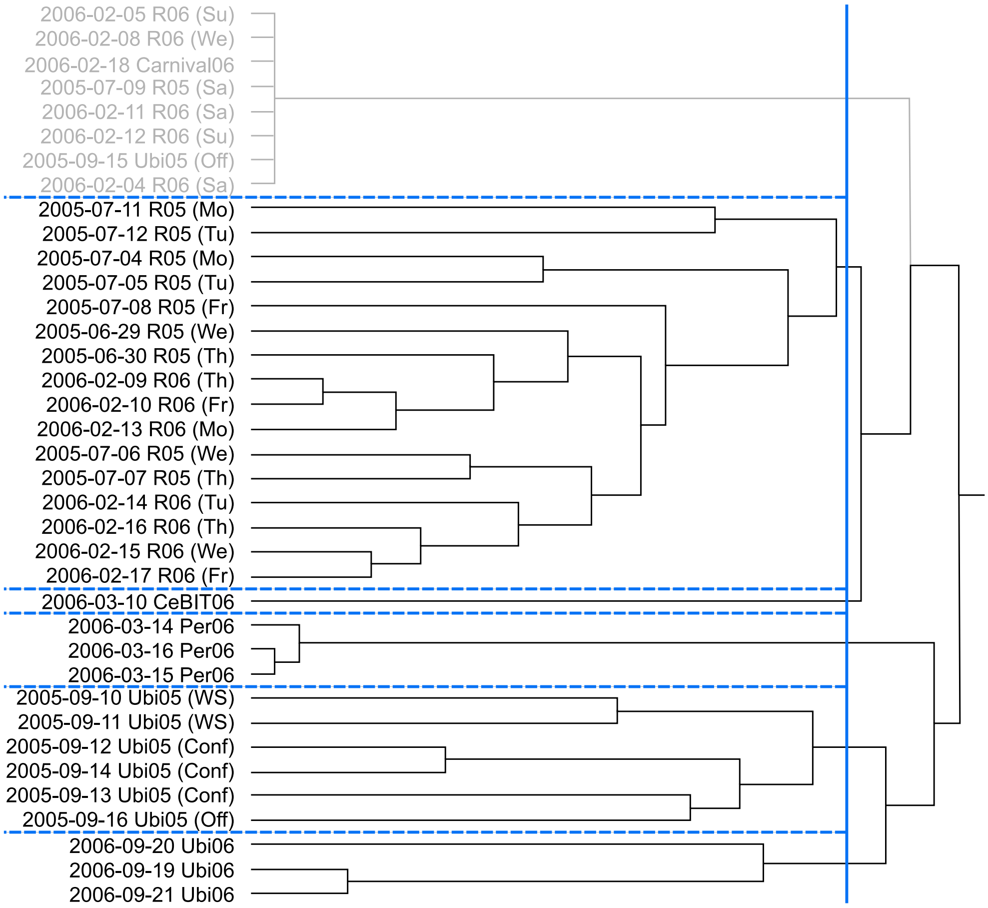


Figure 9.6.: Day clusters, stationary and portable mobility classes

Dataset	Date		Cluster			
			All	W	SP	S
Routine05	2005-06-29	We	1	1	1	1
	2005-06-30	Th	1		1	1
	2005-07-04	Mo	1	1	1	1
	2005-07-05	Tu	1	1	1	1
	2005-07-06	We	1	1	1	1
	2005-07-07	Th	1	1	1	1
	2005-07-08	Fr	1	1	1	1
	2005-07-09	Sa	/	/		
	2005-07-11	Mo	1	1	1	1
	2005-07-12	Tu	1	1	1	1
Ubicomp05	2005-09-10	WS	2	2	2	
	2005-09-11	WS	2	2	2	2
	2005-09-12	Conf	2	2	2	2
	2005-09-13	Conf	2	2	2	2
	2005-09-14	Conf	2	2	2	2
	2005-09-15	Off	2			
	2005-09-16	Off	2	/	2	
Routine06	2006-02-04	Sa	4	4		
	2006-02-05	Su	4	4		
	2006-02-08	We	3	3		
	2006-02-09	Th	3	3	1	/
	2006-02-10	Fr	3	3	1	
	2006-02-11	Sa	3	3		
	2006-02-12	Su	3	3		
	2006-02-13	Mo	3	3	1	
	2006-02-14	Tu	3	3	1	1
	2006-02-15	We	3	3	1	1
	2006-02-16	Th	3	3	1	1
	2006-02-17	Fr	3	3	1	1
Carnival06	2006-02-18	Sa	4	4		
CeBIT06	2006-03-10		3	3	/	/
PerCom06	2006-03-14		5	1	5	5
	2006-03-15		5	1	5	5
	2006-03-16		5	1	5	5
Ubicomp06	2006-09-19		6	2	6	
	2006-09-20		6	2	6	
	2006-09-21		6	2	6	

Table 9.1.: By day dataset overview with clusters ('/' indicates a day in a unique cluster, an empty space indicates, that the cluster could not be determined)

to the workshop days. The two additional days of recreational activities have only weak connections to the other days.

The clustering of all mobility classes (figure 9.3) produced interesting groups. The Ubicomp05, PerCom06 and Ubicomp06 conferences are grouped into separate sets, while the datasets from Germany are mixed. One weekend from the Routine06 data is clustered together with the Carnival06 (also a weekend). Another Saturday (2005-07-09) is also separated from the rest of its dataset. The CeBIT06 on the other hand is combined with the rest of the Routine06 data.

When we take into account the wearable devices only (figure 9.4), the separating effects of time and space diminish. Here, the connections are only formed by people and not by any stationary equipment in the area. Most notably, the two Ubicomp conferences are combined into one cluster, but without the two free days in Tokyo. Also, the PerCom06 is combined with the Routine05 set.

With only the stationary mobility class in contrast, we get a clear separation by geographical location (see figure 9.5). Especially, several days of the Routine06 and Routine05 subsets get combined. Nevertheless, a lot of days do not have enough sightings of stationary devices and must be excluded (grayed out in the figure). When we combine the stationary and portable mobility classes, the picture of the last clustering becomes clearer. There are less days that drop out of the analysis because of missing data.

9.4. Discussion

The subset network in figure 9.1 shows an interesting connection pattern, most of which can be explained easily. The data recorded in Bremen is strongly connected (Routine05, Routine06, Carnival06), since is was recorded in socially and geographically similar situations. Connections between the Routine05/06 and the PerCom06 can be explained, because two colleagues of the proband were on the conference, too. The connections between the Ubicomp05 and Ubicomp06 as well as the Ubicomp06 and PerCom06 are probably because of overlapping attendees. However, an overlap in attendees could not be measured between Ubicomp05 and PerCom06, although there was an overlap. Our proband was able to measure a path from Tokyo via New Beach and Pisa to Bremen and Hanover, which is established by social context.

The day network resembles the same general structure as the dataset network. When we contrast the clusters produced by wearable and stationary devices separately with each other, interesting details about the days in the dataset are revealed. Our assumption here is, that the wearable devices produce groups with socially similar situations, and that the stationary ones produce groups by geographical location. Figure 9.6 supports this assumption for the geographical part, especially as it combines days of Routine05 and Routine06 in one group (with the exception of 2006-02-09). The social groups shown in figure 9.4 are not as clear. The placement of both Ubicomp conferences in one group supports the assumption. The day off (2005-09-16) is consequently not in the same group. The Routine06 subset is combined with Carnival06 and CeBIT06, which makes sense, because there is naturally an overlap in the people seen on the days (they all have at least a few hours in common in the Bremen-context). Interestingly, the Routine05 and Routine06 subsets are strictly separated. The reason for this is probably the gap in time: some people move in, others move out of our context over time. Another reason might be, that the people stayed the same, but that they updated their phones to newer models—thus not being recognized as the same person by our analysis. A peculiarity is the combination of the Routine05 data with the PerCom06. The gap in time is larger, compared to the Routine06 subset, and it took place in another country. An explanation for this could be, that there were old phones involved, possibly for the purpose of traveling.

The differences in geographical location are measured on a city-wide scale: days in Bremen, Hanover, Pisa, Tokyo and Newport Beach are all separated into the corresponding group (see figure 9.6). But our method does not give us details about the different places within those cities. By increasing the number of groups, precision in location does not increase, instead seemingly arbitrary groups form. This is not surprising, because we have chosen a time-granularity of one day. Thus all locations during a day are put together. Instead, we could try the procedure with hour-nodes replacing the day-nodes to determine distinct places within a city.

The selection of the number of groups is a difficult task. The seven we have chosen, and the resulting level of 0.3, seems to be a good choice for this case and yields interesting results, especially when wearable and stationary clusters are contrasted.

Although the eigenbehavior analysis is not particularly expressive on the whole dataset, we could probably use it to analyze the separate groups produced by the method introduced. Since behavior within a group is probably more homogenous, this could produce results which are easier to interpret.

9.5. Summary

In this chapter, we have introduced the temporal-cluster method to visualize temporal entities in the dataset and to compare these entities in terms of their social structure with the goal to form clusters of similar ones. We have demonstrated the method with the separate days of our dataset as well as intervals of several days spanning the subsets of the data. As a result, we have constructed networks of the subsets of data as well as of the separate days. The structures of both networks clearly show, how these temporal entities are connected to each other, based on the social behavior of the proband and his environment.

The clustering of days resulting from the day network based on structural equivalence provides us with concrete groups. Depending on the Bluetooth devices included, we can select to cluster days based on their social context or their location context respectively. With this one technology, we can thus analyze the dataset for its social situation and for geographical location at the same time, by filtering the devices by their mobility class. Location context does not include geographical coordinates in this method, instead locations are represented symbolically. Additional technology, e.g. a database to relate stationary Bluetooth devices to geographic coordinates, is required to map symbolic locations to geographic coordinates.

Regarding the UPI, the temporal-cluster method is able to analyze the aspect of temporal structure on the basis of the aspect of social structure or spatial structure respectively.

10. Social Context of Places

In the previous chapter, we demonstrated how the same data collection method and dataset of Bluetooth proximity data can be used to cluster days by social context and by location context respectively. To switch between social and location context, we use a filter based on our definition of mobility class to select either devices identifying people or devices identifying places.

In this chapter, we present a method to examine social structure and location structure, two important aspects of the UPI, in greater detail. We understand both aspects as being interrelated in the sense of Paulo's and Goodman's idea about *turf* and *tribe* [149]. *Turf* is a public area, or a set of areas, where a person or a group of people feels comfortable in and where he is thus likely to be. *Tribe* on the other hand refers to a group of people with a similar preference for places.[7] It is obvious, that there is an inherent relationship between turf (place) and tribe (people) in the above description. Each concept is characterized by the other one.

Paulos and Goodman have given the concepts of turf and tribe, and they speculated that both could be measured by conducting periodic Bluetooth scans [149]. To characterize the aspects of social structure and spatial structure of the UPI in relation to

[7]By the terms *location* and *place*, we denote similar, but not equal concepts. We use *location* to refer to a specific point in space, which can be identified by a single coordinate. Inspired by [67], we use the term *place* as a human-friendly concept, which usually refers to named areas, e.g. a "marketplace."

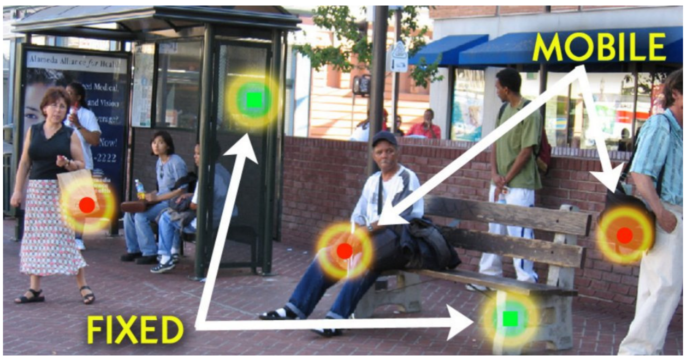


Figure 10.1.: Scenario with mobile sensors attached to people and stationary sensors attached to places [149]

each other, and to continue Paulos' and Goodman's work, we present an appropriate method based on social network analysis.

We demonstrate this method on the subsets #6 (PerCom06), #5 (CeBIT06) and #1 (Routine05) of our dataset of Bluetooth device inquiries (see section 8.2 for an overview about the subsets of the dataset). In #6 and #5 the discovered Bluetooth devices are localized within a building by a number of TrackStations with fixed, known locations (inspired by picture 10.1). The Routine05 subset contains Bluetooth sightings within the city of Bremen conducted by our proband (section 8.2 gives details about the subsets). The discovered devices are localized by concurrent Cell-ID measurements (see section 4.2.1) in this case.

10.1. The Turf-Tribe Method

Our method operates on a combination of social and location information. It requires observations of people visiting certain places recorded in Bluetooth proximity data. Based on this information, we create a network between people and places. Then, we calculate the *social positions* within this network (this concept was introduced in section 3.4.1) to reduce it to meaningful groups of people and places and to quantify the relation between those groups of people and places. The specific steps of the turf-tribe method are the following:

1. A bipartite network of people (or Bluetooth devices in general) and locations is constructed. In this network, a person is connected to a location, if the person was observed to visit that location. This relation is quantified by the number of times this incident was observed. A person must not be connected to another person and a location must not be connected to another location.[8]

2. As an optional step, we lay out the bipartite network on a two dimensional map containing the locations. Depending on the area, a street map, a floor plan or some other map may be used. On this map, the location nodes are placed on the corresponding locations. The person nodes connected to the location nodes can be laid out with embedded spring algorithms (see section 3.4.2). The width of the connecting lines between the nodes may represent the strength of the ties. This visualization gives an impression about the locations in the data and their

[8]Please note: we use a bipartite network for this representation, not a two-mode network as we did in the temporal-cluster method (see section 9.1). The difference is not evident in the network graph, but the matrix representation of a bipartite network is a square matrix. Locations and people are treated equally, in line with each other in the rows and columns of the matrix. Consequently, an operation to transform to one-mode is not required. We choose the bipartite network, because we need to keep both types of nodes in the network for further analysis.

connections by people. However, it can be confusing if there are many nodes and connections.

3. Structural equivalence of the nodes in the network is determined by Pearson correlation (see equation 3.3). As a result, we obtain a matrix containing a similarity metric between all pairs of nodes.

4. Discrete groups of people and locations are then created by single-link hierarchical clustering on the similarity matrix. The output of the hierarchical clustering is a dendrogram, which specifies a progressive separation of the nodes. At this point, we choose the number of groups (people plus locations) we wish to have from the dendrogram. The result is a mapping of each person and location node to a group. To visualize this result, the nodes of the original bipartite network of step two can be marked, e.g. by color and shape, to indicate their group-membership.

5. A blockmodel is formed by combining the bipartite network and the result of the clustering as described in section 3.4.1. In this process, all nodes comprising one group are combined into one node. Because the ties between the original nodes are weighted by the number of visits, the mean value criterion (see section 3.4.1) is used to calculate the weights between the combined nodes of the blockmodel. The result is a blocked network containing groups of people and groups of locations, as opposed to each single location and person in the original network. The number of nodes is equal to the number of groups chosen in step four.

6. Finally, we lay out the blocked network in the same way as the bipartite network in step two. Additionally, the size of the nodes corresponds to the number of original nodes a blocked node represents. In contrast to the network of step two, the blocked location nodes may comprise a set of locations. This can either be visualized by placing the blocked location node in the center of the separate locations it represents or by drawing the boundaries explicitly. The resulting figure shows the different groups of people, groups of locations and the relations between them in a clear picture. By choosing different numbers of groups in step four, the granularity of groups can be adjusted.

The following three sections demonstrate this method on three different subsets of our data. The method is varied slightly, if necessary. We also show how the resulting data can be visualized.

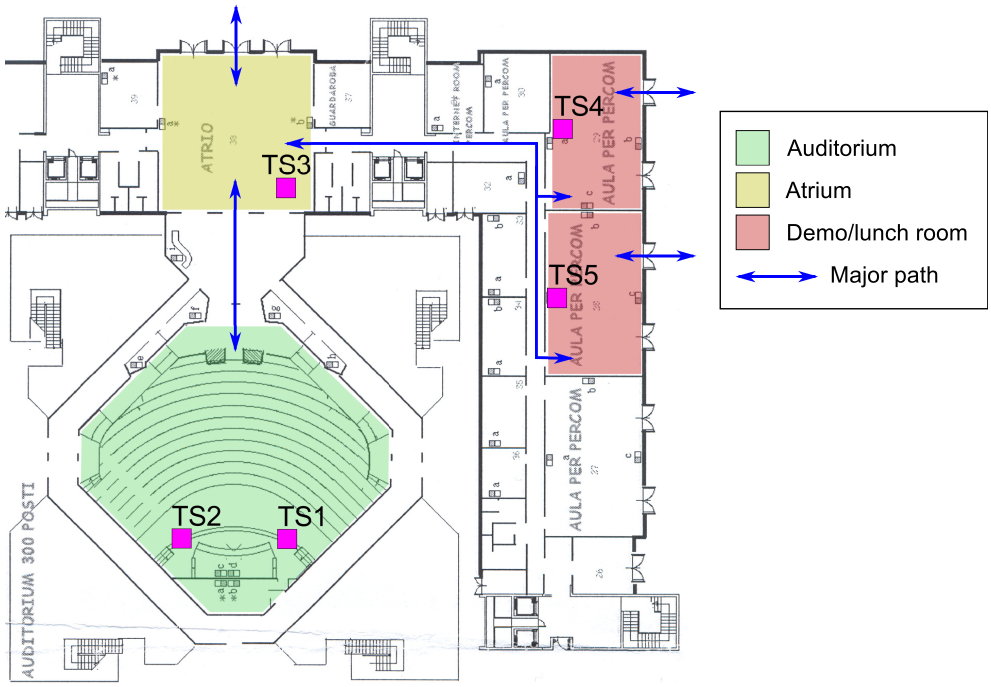


Figure 10.2.: Locations of TrackStations in the PerCom06 experiment

10.2. PerCom06 Experiment

We introduced the PerCom06 subset in section 8.2. So far, we limited the analysis to the data from our proband's mobile scanner. In this section, we analyze the complementary data, which was recorded by stationary scanners at the same conference.

The PerCom 2006 conference was hosted in the CNR research area. It was a single-track conference with all paper presentations taking place in one auditorium. Demos and posters were located in the same rooms where lunch was served. The main entrance to the location was in the atrium, but people could also enter the building through the demo/lunch rooms. The registration was located in the atrium, too, and it connected the auditorium with the demo/lunch rooms. During the three main days of the conference, we placed five TrackStations within these premises. Their placement is depicted in figure 10.2.

During the conference days, 67 devices were discovered by the TrackStations. To set the devices into relation with the locations where they were discovered, we built a bipartite network with the TrackStations connected to the devices they discovered. The strengths of the ties correspond to the number of sightings. Figure 10.3 shows the resulting network. The squares represent the TrackStations, circles are discovered

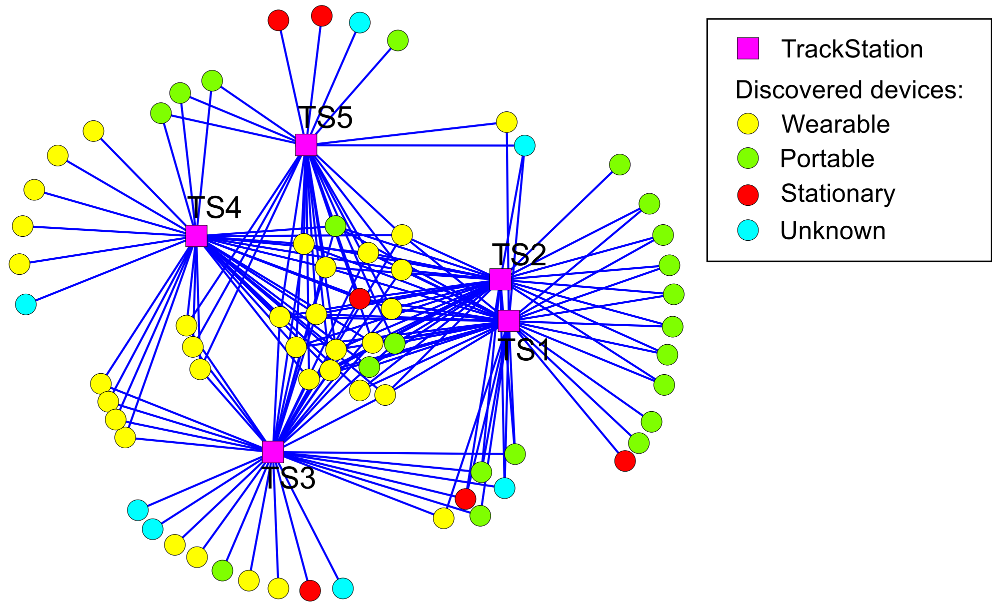


Figure 10.3.: Network of devices at PerCom06, colors indicate mobility classes

devices. The colors indicate their mobility classes (see section 4.5 for the mobility class definition).

Next, we determine structural equivalence of the nodes in the matrix with correlation, as we did with the day network in section 9.3. Hierarchical clustering is applied and a clustering with six groups is chosen. Figure 10.4 shows the resulting groups. TrackStations (squares) and discovered devices (circles) are located in different groups.

As a final step, we create a blockmodel from the clustering (see section 3.4.1). Because our data contains weighted ties quantified by the number of sightings, we use the mean value criterion to determine the strengths of the ties in the blocked network. The blocked network is shown in figure 10.5. Matrix A shows the corresponding matrix, the numbers of devices v are used for the sizes of the nodes in the network. The number labels of the nodes in the blocked network figure correspond to the columns and rows in matrix A and vector v.

$$A = \begin{bmatrix} 0 & 0 & 0 & 15.3 & 3.8 & 1.1 \\ 0 & 0 & 0 & 3.8 & 21.6 & 0 \\ 0 & 0 & 0 & 1.3 & 3.7 & 7.0 \\ 15.3 & 3.8 & 1.3 & 0 & 0 & 0 \\ 3.8 & 21.6 & 3.7 & 0 & 0 & 0 \\ 1.1 & 0 & 7.0 & 0 & 0 & 0 \end{bmatrix}$$

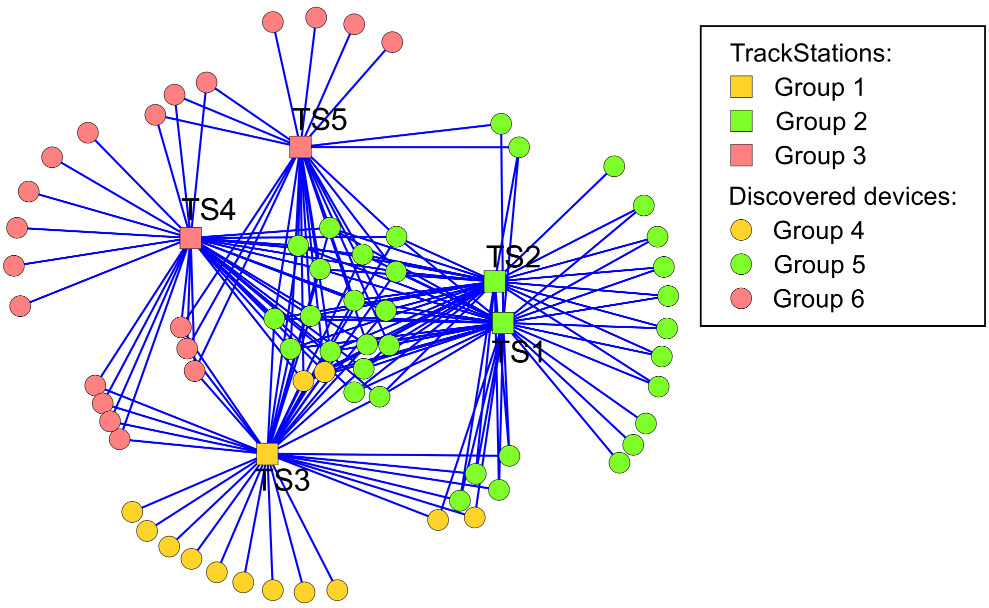


Figure 10.4.: PerCom06 network with six groups by colors and shapes

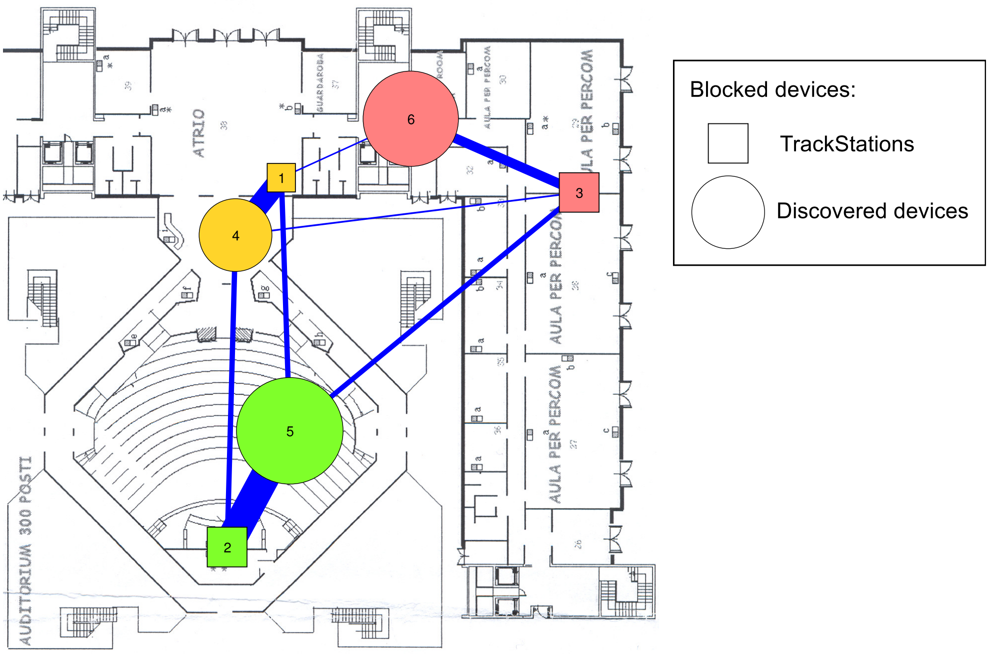


Figure 10.5.: Blocked network of PerCom06

$$v = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 7 \\ 15 \\ 12 \end{bmatrix}$$

The blockmodel of the PerCom06 experiment (figure 10.5) gives a very clear overview about the different groups of Bluetooth devices at the conference, based on their pattern of movement through the rooms. It also combines the stationary scanners in a meaningful way: the two scanners in the auditorium were combined into one node, as well as the scanners in the demo/lunch rooms. The scanners were placed so that conference attendants would spend a fair amount of time at all their locations. Nevertheless, several devices were discovered by only a subset of the scanners, especially those represented by node six. In the auditorium, there is a noticeable accumulation of portable devices (see figure 10.3).

With knowledge of the general patterns of behavior at the observed event and the spatial layout of the blockmodel on top of the floor plan (figure 10.5), we can interpret the different groups. Node five represents a pattern of behavior that we would expect of visitors to the conference. They spent most of their time in the auditorium to listen to the talks. The accumulation of portable devices indicates that people were using their laptops in the auditorium (see figure 10.3). This group is also connected to the entrance area and the demo/lunch rooms. Node four represents a pattern explained by the behavior of administrative staff. There is a strong connection to the entrance room, where the registration and the info counter were located. Node six in contrast shows devices that were not used in the auditorium at all. It is strongly connected to the demo/lunch rooms and we expect this to be caused by catering staff and people mainly concerned with the demos.

10.3. CeBIT06 Experiment

We used a similar setup, as we did at the PerCom06, at the CeBIT 2006, an annual major computer trade fair taking place in Hanover, Germany. As a variation, we mixed stationary and mobile scanners in this example. Two TrackStations were set up at different booths in different halls. One was located at the Bremen stand, the other at the Microsoft stand. Our proband was visiting the fair with a mobile scanner.

Besides a lot of stationary devices at the fair, 673 wearables were discovered by the three scanners, which were used for the further analysis. Figure 10.6 shows the three scanners and the discovered wearable devices as a network diagram. The network is not bipartite as the last one, where connections between the scanning devices were ignored. Instead, the proband was visiting both stands, which resulted in mutual sightings between the proband's device and each of the stationary scanners.

Next, we repeated the application of the turf-tribe method, as we did in the previous section. Structural equivalence was measured with correlation, groups were formed with hierarchical clustering, and a blockmodel was constructed. Ties with a value less than 1 were removed to exclude very weak connections. The resulting network is shown in figure 10.7. The corresponding matrix A quantifies the strengths of the connections. Vector v, containing the number of devices in each node, was used for the sizes of the nodes. The number labels of the nodes in the blocked network figure correspond to the columns and rows in matrix A and vector v.

$$A = \begin{bmatrix} 0 & 14 & 0 & 1.14 & 0.08 & 1 & 2.24 & 0 & 0.46 & 0 & 0 \\ 14 & 0 & 18 & 0 & 9.46 & 1.67 & 0.06 & 6.1 & 7.35 & 0 & 2.66 \\ 0 & 18 & 0 & 1.14 & 3.69 & 0 & 0 & 18.33 & 7.04 & 1.66 & 0.05 \\ 1.14 & 0 & 1.14 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.08 & 9.46 & 3.69 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1.67 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.24 & 0.06 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6.1 & 18.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.46 & 7.35 & 7.04 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.66 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.66 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$v = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 7 \\ 13 \\ 9 \\ 143 \\ 21 \\ 26 \\ 108 \\ 346 \end{bmatrix}$$

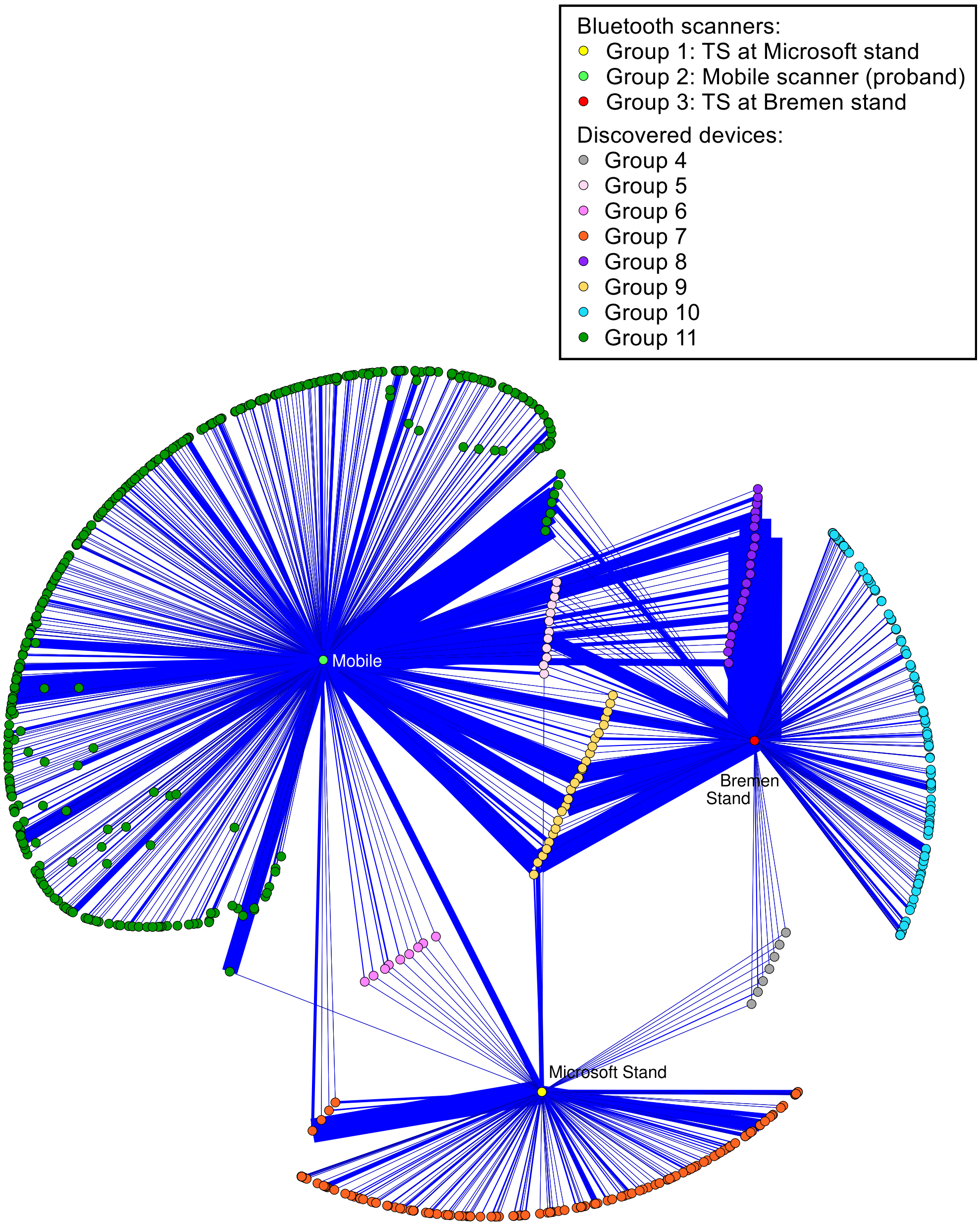


Figure 10.6.: CeBIT network, colors indicate groups

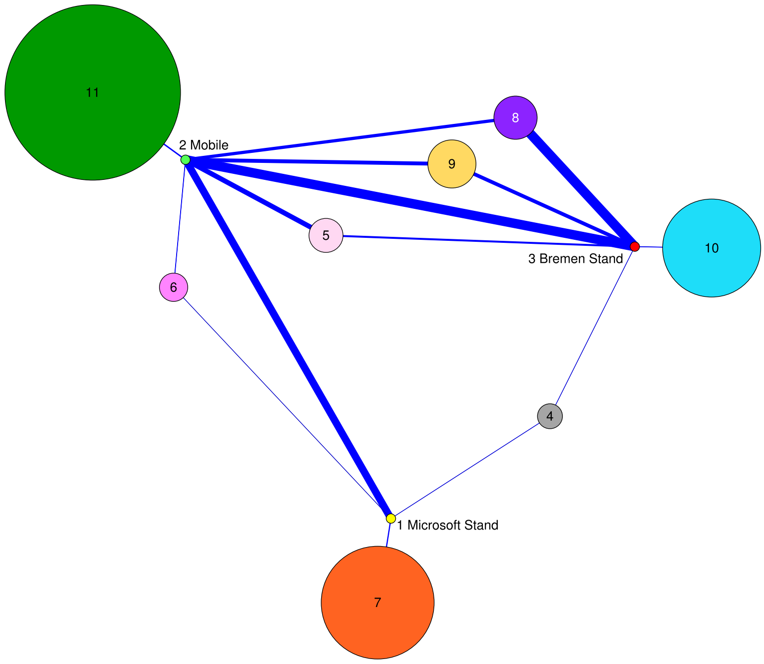


Figure 10.7.: Blocked CeBIT network, edges with values less than 1.0 removed

While the amount of people was limited on the PerCom conference, there was a massive amount at the CeBIT: 673 wearable devices. With our focus on wearable devices in this experiment, we can relate devices to people. The blocked network of figure 10.7 shows, that the largest group of people (node 11) was discovered by the proband with the mobile scanner (node 2). He visited the Bremen stand (node 3) for a longer period of time than the Microsoft stand (node 1). This behavior is consistent with the data, showing that there were more different devices discovered at the Bremen stand (nodes 4, 5, 8, 9 and 10), than at the Microsoft stand (nodes 4, 6 and 7).

Further, the blockmodel provides us with groups of people who were with the proband at the stands. It differentiates between people being discovered more often at the stand than by the proband (node 8), that were discovered more often by the proband than at the stand (node 5 and 6) and groups being discovered by the proband approximately as often as at a stand (node 9). We also get a small group that was discovered at both stands, but not by the proband (node 4).

10.4. Routine05 Experiment

Next, we turn to another variation on the connection between social context and location. The Routine05 subset contains Cell-ID measurements additional to the Bluetooth discoveries. This gives us the locations of the proband within the city with an accuracy of approximately 100–200 meters. Further, we can locate the Bluetooth contacts in space. Compared to the stationary Bluetooth scanners used before, there is no need to equip each place with such a device. The drawback of this method is, that the resulting location information is less precise (see section 4.2.1).

A bipartite network is constructed from the data. Cell-IDs, as well as wearable Bluetooth devices, are both taken as nodes. A Cell-ID node is connected to a Bluetooth device node, if the proband's mobile scanner detected a device, while it was logged into a certain cell. The strength of a tie corresponds to the number of sightings within a cell. The proband's device is not explicitly present in the constructed network, but implicitly connected to all nodes.

After the dataset has been collected, the connection between Cell-IDs and real-world locations had to be established. Based on a list of places the proband reported he had been, those locations were explored for the present cells. Then, the network could be laid out in a geographical manner by tying the Cell-ID nodes to locations on the map (see figure 10.8).

With the same procedure as in the last experiments, structural equivalence is determined, groups are created and a blockmodel is formed. In figure 10.8, the groups of this procedure are indicated, which are then reduced to the blocks of figure 10.9. The corresponding matrix A and vector v show the underlying numerical values. Again, the number labels of the nodes in the blocked network figure correspond to the columns and rows in matrix A and vector v.

$$A = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.33 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.25 & 0.25 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.03 & 1.87 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.17 & 0 & 0 & 0 & 0.5 & 0 & 1.5 \\
0 & 0 & 0 & 0 & 0 & 0 & 9.39 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 171.77 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 9.39 & 171.77 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1.33 & 0 & 0 & 0.17 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1.25 & 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.25 & 1.87 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

$$v = \begin{bmatrix} 1 \\ 2 \\ 15 \\ 2 \\ 63 \\ 11 \\ 2 \\ 3 \\ 2 \\ 2 \\ 2 \\ 1 \\ 2 \\ 2 \end{bmatrix}$$

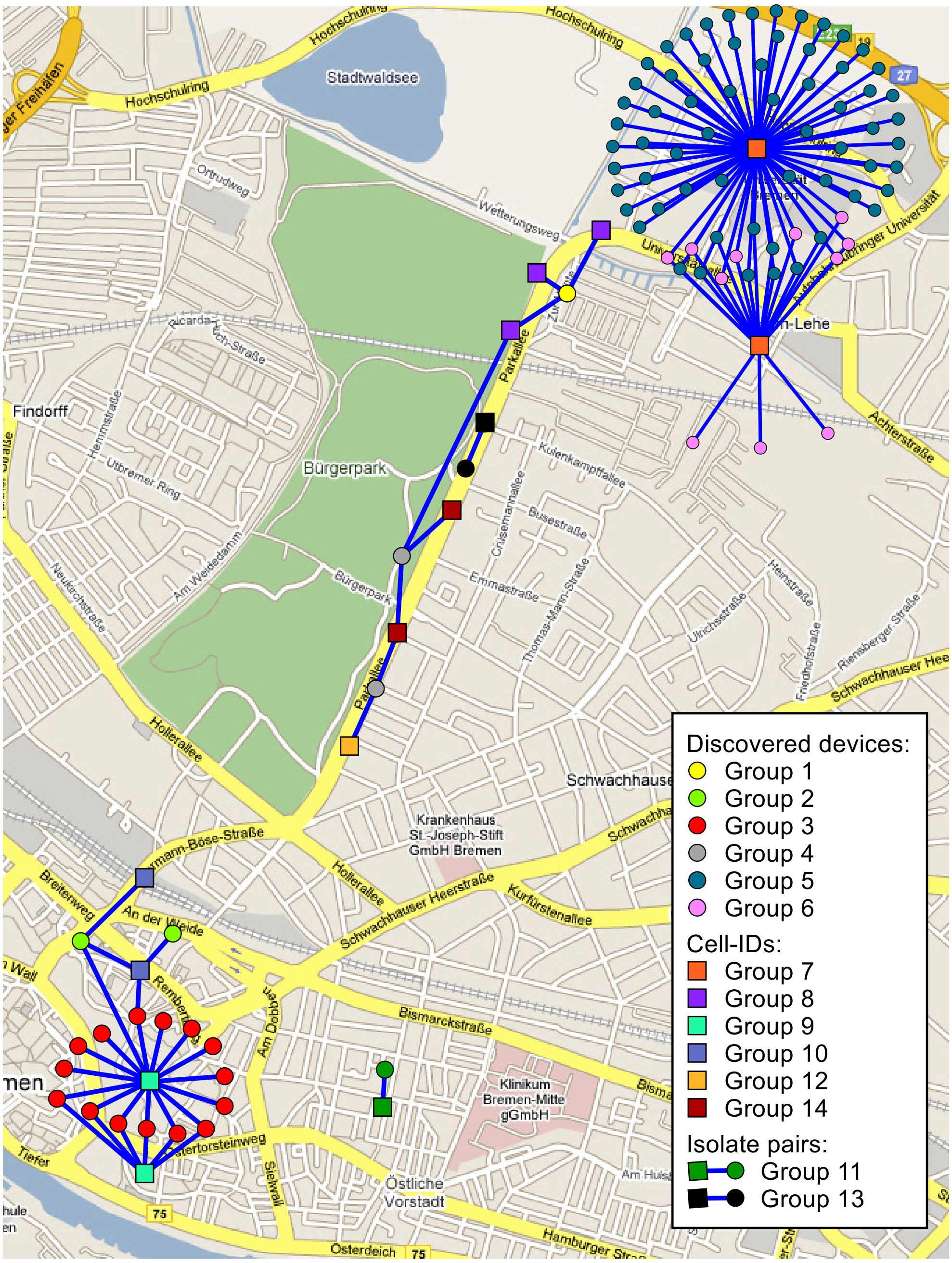


Figure 10.8.: Bipartite network of Bluetooth devices and Cell-IDs, located on a street map

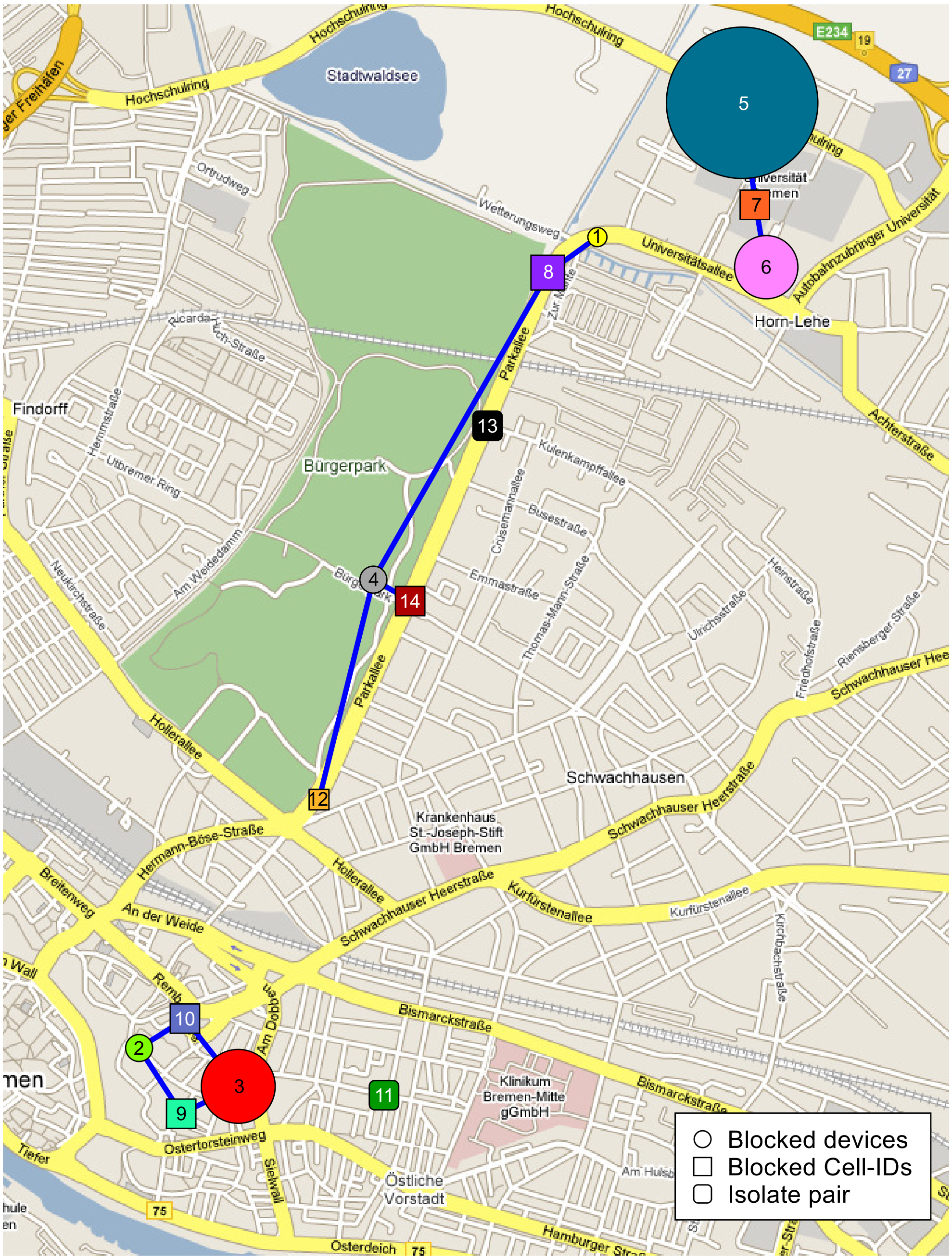


Figure 10.9.: Blocked network of Bluetooth devices and Cell-IDs, located on a street map

The Cell-ID location method of the third experiment and the resulting blockmodel gives an overview about the locations our proband has been, in combination with people he discovered (figure 10.8). In the north-east, there is the university, where he was working. The city center in the south-west is where his apartment was located. He traveled the way between home and work by car, along the park. The network consists of five disjoint components. The major components of the network represent three basic situations: home, work and commute. The two places, home and work, are socially unconnected in this example. There is no overlap in discovered Bluetooth devices at those two places. Further, there are two isolate pairs, single discovered devices connected to single Cell-ID nodes.

The blockmodel is again able to simplify the network (figure 10.9). In comparison to the other two experiments, we have a larger number of locations here. Twelve different Cell-IDs—excluding the isolate pairs—get summarized to six in the blockmodel. The reduction by structural equivalence does not work correctly with the two isolate pairs of Cell-ID node and device node. In the blocked network, they are reduced into one cluster, in contrast to the other cases, where the distinction between Cell-ID node and device node is preserved. Thus, it would make sense to separate such isolate pairs and exclude them from the procedure.

10.5. Discussion

With the given examples, we could show that the turf-tribe method presented in this chapter is able to create discrete groups of discovered people by their preferences for certain locations. This works in a variety of different settings: while the data at the PerCom06 is rather dense, both in terms of location and sightings per Bluetooth device, it is sparse in the Routine05 set. The CeBIT06 data again exhibits a different pattern: there is a large amount of detected devices, but the recorded location information is rather sparse which limits the interpretation of the groups of people.

In a similar manner as people are combined into groups, different locations are combined, if they exhibit a similar social context. This is demonstrated with the PerCom06 data (the two locations in the auditorium as well as the two demo/lunch rooms) and with the Routine05 data (we can see this reduction throughout the components: home, commute and workplace). This means, that the method achieves a differentiation of locations by their surrounding social context. Although there is no example in our dataset, we can also expect, that geographically distant locations are combined, if they are frequently visited by the same persons. Especially on a large scale dataset comprising a whole city, such a network visualization could reveal interesting social connections between different neighborhoods.

Another aspect of the turf-tribe method is that it can cope with different perspectives of observation. The PerCom06 experiment demonstrates an allocentric perspective, with a number of stationary scanners overlooking specific locations of an area. In the Routine05 experiment in contrast, the whole analysis is subject to the egocentric perspective of the proband. Thus, the result does not represent the real connections between people and places, it is rather filtered through the personal view—it represents the observations of the proband. The CeBIT06 experiment shows the application of the method to a mixed scenario: there are two locations and the mobile proband. The outcome must thus be interpreted in the light of both allocentric and egocentric views. Though mobile and fixed scanners are not partitioned in the structure of the matrix, they stay distinct for their patterns in the CeBIT06 experiment.

We have also used different localization techniques and applied the turf-tribe method to them. On the one hand, there is the high-accuracy technique of Bluetooth localization (PerCom06 and CeBIT06), on the other hand, there is the rough Cell-ID approach (Routine05). The experiments in this chapter show, that our method can manage both equally well. It would be interesting to test the turf-tribe method on more different localization techniques. E.g., GPS would result in a much larger number of locations—every single localization would result in a slightly different location. Because our method is based on the structural equivalence of the relation between locations and people, it should be able to create meaningful clusters of GPS locations. Another interesting approach would be to combine multiple localization techniques (e.g. Bluetooth, GPS and Cell-ID).

A variation we have not explicitly looked at is the visualization in different resolutions. In the examples, we have chosen specific numbers of clusters, and thus the number of groups that should be created for the blockmodel. By changing this number, it is possible to create a detailed view with many nodes (the most detailed view does not cluster at all and is thus shown in figures 10.4, 10.6 and 10.8) or an overview with meaningful groups (e.g. figures 10.5, 10.7 and 10.9).

10.6. Summary

In this chapter, we exploited geographical location information to localize social context in space and to construct geographical areas of similar social context. We introduced our turf-tribe method and demonstrated it on three different subsets of our dataset exhibiting different situations. The evaluation shows that the method is able to analyze data recorded from an egocentric perspective, the allocentric perspective or mixed points of view. The resulting network graphs of the method can be visualized geographically, e.g. on a map or a floor plan. Social structure can thus be explored in a convenient way.

When we compare our findings to the notion of "comfort in public places," as put forward by Paulos et al. [149], we see that the method in this chapter can visualize the concepts of *turf* and *tribe* in a novel way. Moreover, it contributes the composition of social groups as well as the construction of larger areas from multiple location measurements.

The turf-tribe method builds on two important aspects of the UPI—social structure and spatial structure—and allows their interpretation in relation to each other.

11. Detection of Episodes

In chapter 9, we presented a method to analyze the dataset on a rough scale of time. From the aggregation of the whole dataset or subsets of a couple of days, we proceeded with a day to day analysis. In this chapter, we conclude the analysis of our dataset with a view on social context on a smaller scale of time and analyze personal behavior within a day.

The focus is on the changing of the social environment in five minute steps, as a way to separate episodes during the day. The basic idea is, that many episodes are separated by a change in social context, or are characterized by a sustained change of it. Therefore, we propose two feature functions which indicate a changing in the social environment, both within that of familiar Bluetooth devices and strange ones respectively. The features are chosen to be independent of the percentage of people that can be identified by the device inquiries, but to achieve good performance, enough detectable devices are a requirement.

The method is applied to the Ubicomp05 subset. We found, that we can easily match the results of our feature functions to the schedule of the conference and recognize deviations, where the proband did not follow the schedule.

11.1. Procedure

We base the detection of personal episodes on Bluetooth proximity data collected from the egocentric perspective. For the intended analysis, data is necessary in a relatively high resolution (approximately 30 seconds in our case). It is also important to have continuous data covering the whole time of measurement, preferably without any dropouts.

The Ubicomp conference 2005 in Tokyo together with the workshop "Metapolis and Urban Life" was selected as a social event for the experiment for its varied program schedule, and because it was expected that a large proportion of the conference attendees had a detectable Bluetooth device with them. Our proband was one of the attendees, carrying a mobile phone with the WirelessRope J2ME program during the entire time of the conference, including time he spent before and after the official conference time schedule outside of his hotel room. Additionally, he took photographs with the same device to document his activities. The program schedule of the confer-

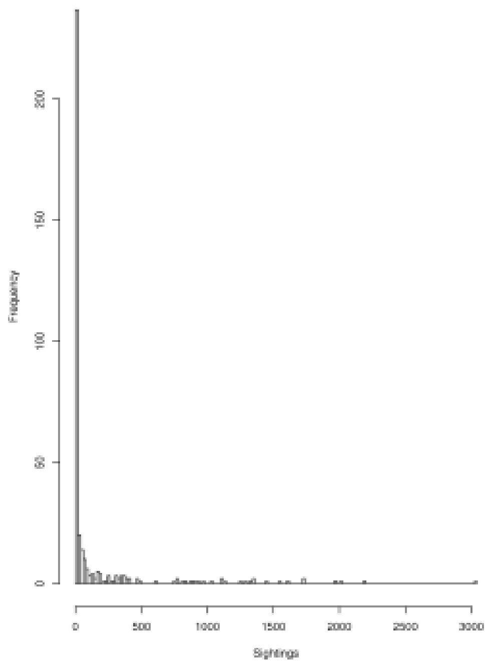


Figure 11.1.: Histogram of sightings

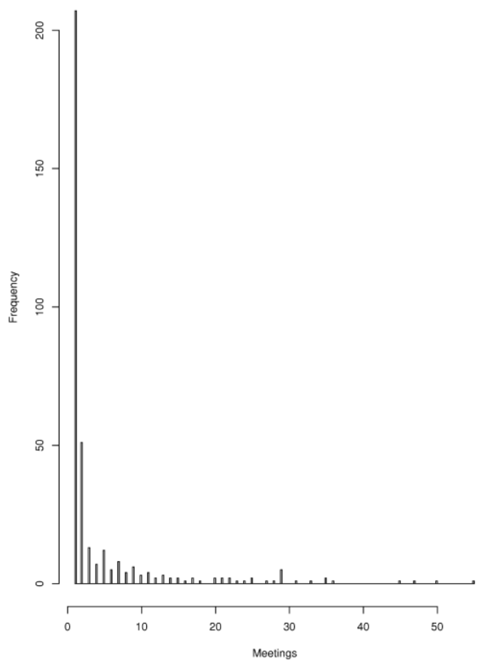


Figure 11.2.: Histogram of meetings

ence provides detailed information about the planned timing of activities. The WirelessRope was configured to conduct scans every 30 seconds to achieve high resolution data. The experiment ran over six days in September 2005. On day one and two, the workshop took place. Part of the first day was an exploration of the city in the afternoon. Days three to five were spent on the main conference. The last day was spent with recreational activities in the city.

The resulting dataset comprises 50,953 Bluetooth sightings (compare to table 8.1). From the recorded dataset, only the wearable device classes were used for further analysis (see section 4.5), which have the highest correlation with people. These devices were classified into either *familiar* or *strange*, following a basic bipartite view of people from an individual's perspective. The model we use to accomplish this classification is straight forward and works reasonably well in the conference scenario. We define a *familiar device* to be one, that was met more than five times by the proband. A *meeting* is defined as a duration of time, during which the device was regularly detected. After five minutes of absence, a meeting was over. This method regards for common failures of the Bluetooth inquiry mechanism to detect a device from time to time. Following this method, there are 1,661 meetings in total in the dataset. Figure 11.1 and 11.2 show the histograms of individual Bluetooth sightings and the derived meetings, respectively. There were approximately 650 registered conference visitors.

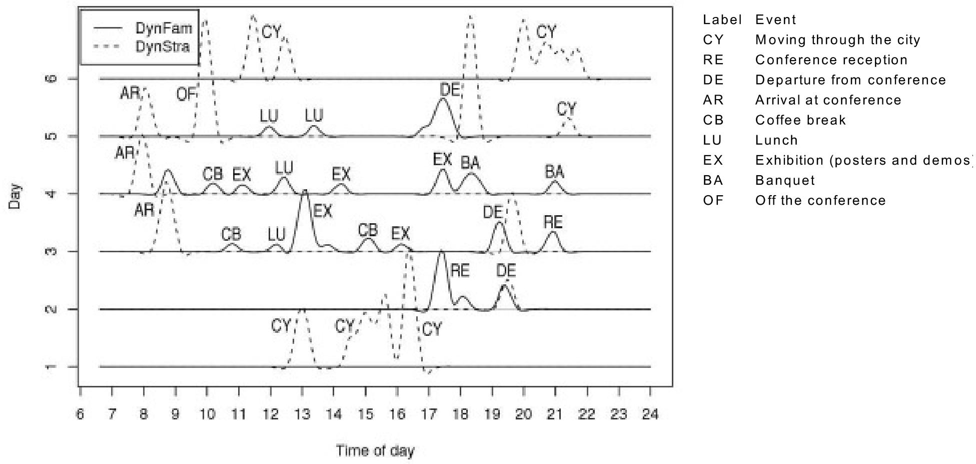


Figure 11.3.: Feature data of six days in Tokyo (smoothed by splines)

69 devices were classified as familiar and a total of 290 as strangers for the whole data set including conference and city encounters.

For the analysis, a set of quantitative features was extracted from the sets of devices by a sliding time window of five minutes. Let F_t be the set of all detected familiar devices in the time interval $[t, t+1[$ and S_t the set of strange ones, respectively. The number of arriving familiar devices is $f_t^+ = |F_t| - |F_t \cap F_{t-1}|$ and $f_t^- = |F_{t-1}| - |F_t \cap F_{t-1}|$ is the number of leaving familiar devices. s_t^+ and s_t^- are defined correspondingly. The analyzed features indicate the dynamic in the group of familiars and strangers. They show how much an individual moves in accordance with the surrounding people:

$$DynFam(t) = \frac{(f_t^+ + f_t^-) - ||F_t| - |F_{t-1}||}{|F_t|} \tag{11.1}$$

$$DynStra(t) = \frac{(s_t^+ + s_t^-) - ||S_t| - |S_{t-1}||}{|S_t|} \tag{11.2}$$

11.2. Results

Figure 11.3 shows the features DynFam and DynStra for the six days of the experiment. Labels assign meaning to the events detected by the feature functions, e.g. arriving at the conference, coffee breaks and time spent at the exhibition and poster session. The peaks indicate the different social activities the proband was engaged in. The conference activity shows up clearly in the data. Arrival is indicated by a peak in DynStra that is triggered during the movement through the crowded city. Coffee breaks, lunch and visits to the exhibition are indicated by peaks in DynFam. The workshop during day one and two is not detected, since the group behavior was rather homogeneous and did not exhibit the measured dynamic. The city exploration as part of the workshop on the other hand is clearly indicated. The arrival to the workshop did not require movement through crowds.

In detail, the diagram can be interpreted for the following situations on day three:

7:45 – 9:00 The proband was going from his hotel to the conference venue by subway. The data shows an increase in strange persons, then the passing by of strangers and subsequently a decrease of strangers and an increase of familiars as he arrives at the conference.

9:00 – 18:50 There are five peaks in the dynamic of familiars during this period. They correlate with the conference program and indicate situations of moving among the familiar conference attendees. The first coffee break was at 10:40, lunch began at 12:00, during the third wide peak the test subject was talking to different people at demo stands. The next coffee break was at 15:00. The demo and poster session began at 16:40.

18:50 – 19:30 Attendees were transferred to the reception location by buses. Dynamic in familiars is due to the departure situation, with an increase of strangers and decrease of familiars when people were separated. Passing by of strangers is indicated by the peak in the dynamic of strangers. The increase in familiars shows the arrival at the reception.

19:30 – 21:30 The proband was at the reception. Dynamic in familiars reveals that he was changing the group he was talking to at 20:45.

21:30 – 0:00 After the reception, he went to a restaurant with some of the conference attendees. The data shows a decrease of familiars at departure, but the restaurant situation is not visible in the data. At 23:15 he left the restaurant and returned back to his hotel.

11.3. Discussion

The peaks vary in width and height. The height relates to the frequency and amount of the changing of people in the surrounding and the width to the duration of the changing. Thus, the departure from the conference with everybody leaving at the same time causes a higher peak than the coffee breaks, where only a subset of the attendees has been in motion. The whole conference situation is characterized by the paper sessions with regular breaks for coffee, etc., resulting in short peaks on the feature functions. Continuous movement through the crowded city, in contrast, produces an amplitude over a longer period of time.

Our model of partitioning familiar and strange Bluetooth devices works well for the given scenario. Movement in the anonymous city can be clearly distinguished from movement in the familiar context of the conference. With the knowledge of the larger context—the conference visit in this case—it is possible to assign meanings to the individual peaks. We can correlate scheduled events with the characteristics of the feature functions. We can also recognize deviations from the schedule, e.g. the short-term departure of the proband from the conference on day five, at 10:00, from the peak in DynStra.

There were a few limitations encountered with this experiment. First, Bluetooth was generally unpopular in Japan. However, most of the times there was enough reception in the city for this analysis. Only movement in the night was generally not detected, although there were strangers on the streets. Inaccuracies in Bluetooth device inquiry were also discovered, but seem to have no significant negative effect (compare [46]). Moreover, the processing could not have been carried out in this manner during the measurement in a real-time fashion. The reason is, that the familiarity was calculated over the whole conference time before the features were calculated. Thus, effects of the process of becoming familiar are not addressed here.

The approach presented in this chapter is fundamentally different from the methods of the previous chapters in that it is not based on network structure. Instead, it follows in the tradition of established machine learning methodology, where the behavior of a system is understood in terms of the combination of a set of feature functions. The presented functions DynFam and DynStra can be used in such a methodology to identify and separate different episodes in daily life, which are characterized by social context. In combination with a more sophisticated algorithm to partition Bluetooth devices into more than the familiar/stranger dichotomy, e.g. by determining the “tribe” of a person based on the turf-tribe method of chapter 10, a powerful classification engine could be built.

11.4. Summary

In this chapter, we have used high resolution Bluetooth scan data to recognize personal episodes during a day. Again, our method relies on social context to facilitate the analysis, but unlike the previous chapters, it does not utilize network models.

A simple function to partition Bluetooth devices into familiar and strange devices on the basis of meetings was introduced. On this basis, two feature functions were defined to measure dynamics in the Bluetooth environment—one for the dynamic in familiar, the other for the dynamic in strange devices. The functions' amplitudes relate to the changing of the social context in five minute steps throughout the day.

The application of these functions to the Ubicomp05 subset have shown, that it is easy to correlate the peaks and amplitudes with the schedule of the conference. Deviations from the schedule can be identified and the patterns generated by recreational activity in the city exhibit a very different characteristic.

Part IV.

Conclusions and Future Work

12. Conclusions and Future Work

In the beginning of the 2000s, I was showing a group of teenage school girls the wearable computing lab in the TZI, Bremen, with all its fancy and futuristic technology. To make an argument, that—at least a couple of them, I expected—were already having some kind of computer with them, being always on and connected, I asked them, who of them owned a mobile phone. The reaction was not quite what I anticipated: they looked at me, as if I were from out of this world, and told me, that of course they all had one. At that time, I myself did not own a mobile—I definitely was out of their world.

Today, the mobile phone is a pervasive technology, not only for the digital natives—I also have one (my third so far). Recently, these devices evolved into a platform for mobile sensing [23] and feature powerful interfaces with the ability to browse and manipulate impressive amounts of data. Over night, it seems, research visions have been turned into commercial applications (e.g. augmented reality). But still, there are numerous opportunities for the phone to fill the gap between digital and sensor city, as we have outlined in chapter 2. How can we build upon the current infrastructure to reach the next stage to reveal a myriad of opportunities for cooperation, innate to our urban environments?

We have undertaken the attempt of understanding a few facets of what makes up temporary urban life, especially the ambiguous role of anonymity and (familiar) strangers, which proves to be defining of our urban experience. Chapter 3 has further identified current usages of social context in pervasive computing applications. There is a trend for such applications to leave the controlled office environment they were initiated in and move out onto the streets—supporting us in making sense of the complex social structure surrounding us.

From a pervasive computing perspective, measuring spatial proximity is key to the understanding of social context. When it comes to spatial proximity, the pervasive sensing platform mentioned in the beginning of this section with its Bluetooth short range communication protocol bears tremendous potential, compared to other technologies (see chapter 4).

Most users underestimate the amount of traces they leave behind when using networked information technology in general as well as the insights and conclusions that can be drawn out of them (e.g. by the advertising industry). Mobile technology is no different in this respect and is eventually even able to provide more—being with the user all the time, being equipped with all kinds of sensors, becoming a general interface to all kinds of services. While there are traces which are an inevitable necessity for the functioning of the technology itself (e.g. the traces of cell tower connections recorded by the operators of mobile networks), people voluntarily provide more than this: we found, that depending on the location and cultural background, there are approximately between 2% of people in Germany, 6% in the US and 7% in the UK (measured in 2006, see chapter 7), which can be detected by conducting standard Bluetooth device inquiries.

A lot of people are aware of the visibility of their phones as we can see when taking a look at all the funny and expressive names they have given to their devices. Some are even requests to interact with them and to send them messages [89]. The component which turns Bluetooth into an excellent tool for studying social context is the unique and immutable device address of each device. You might ask yourself: where have I encountered that person over there before? Your mobile can measure and record the different social circles you visit and give you an exact answer to your question. It could also tell you other places a person tends to go. Maybe you should check them out?

The premise of our work is, that the few percent of people we can discover using Bluetooth technology are representative for the greater society. Thus, this technology gives rise to measure human mobility and social circles at a precision, which is by far higher than the pure Cell-ID data recorded by mobile operators. Although this might seem to have Orwellian aspects of surveillance, the Bluetooth approach is actually closer to the concept of sousveillance, as Mann puts it forward [116], enabling everybody to collect and use this freely available information for his own purposes.

To conduct Bluetooth measurements on a large scale, we have developed the WirelessRope proximity sensing system consisting of mobile scanners, stationary easy-to-deploy scanners and data collectors as well as a server to aggregate all data and transform it in different ways. All the necessary technology we built upon is widely available. Following the philosophy of sousveillance, we put the developed software in the public domain, for everybody to learn about his own social context (see chapter 5). Our proband was using the equipment in seven different settings ranging from daily routine to conference situations, sightseeing and a carnival procession.

12.1. Contributions

In chapter 2, we presented the concept of the *urban pervasive infrastructure (UPI)* with its aspects of mobility, temporal structure, social structure, spatial structure and facts and figures. Throughout this thesis, we substantiated the concept with the application of augmented gatecounts, methods to put social, spatial and temporal structure into relation as well as a system to measure Bluetooth signals of the UPI. The principal contributions are:

Visualization of Social Context We have developed a method for the mapping and visualization of social context. The method is based on the construction of an ego proximity network based on physical proximity. By measuring the Bluetooth devices around us, we can locate any social situation on the map and put it in relation to our former experience.

We have demonstrated this method with data collected by our proband through the course of 37 days. We have shown how to construct an ego proximity network from our dataset. The notion of 2nd grade encounters is used, where a relation between two nodes in the network is created when two devices are encountered at the same time and location.

Our data demonstrates, that such a network is highly connected with the majority of the nodes in one component. Thus, we could show that even the small percentage of people we could measure with our method are enough to create a mostly cohesive map of personal social context. Chapter 8 contains a detailed discussion.

Comparison of Days by Social Context Further, we have developed a method to compare temporal entities, e.g. days or sets of days, in terms of similarity of social context. the metric gives us the possibility to cluster days we have spent in socially similar situations. Thus, we can create groups of days, e.g. for the purpose of creating an automatic diary.

We facilitated this by first creating an affiliation network of the temporal entities in question and the discovered devices. Structural equivalence proves to be an apt measure to perform the final hierarchical clustering.

Interestingly, by filtering for different Bluetooth mobility classes (e.g. wearable or stationary, as defined in section 4.5), we can shift the focus from similarity in social context to similarity in geographical location. Chapter 9 shows how the method performs on our dataset.

Comparison of Locations by Social Context With a variation of the previous method, we are able to identify and visualize local communities of people as well as locations these communities are related to. By using an additional localization method (e.g. Cell-ID or with a mapping to localize stationary Bluetooth devices geographically), we can determine the geographic locations the local communities are connected to. We can also summarize geographic locations in terms of their social context. E.g., some streets in a city might be inhabited by similar people, going to work and recreation at the same places, while a neighboring street might be inhabited by another class of people. Such segregation can be observed in many cities.

For this method, a bipartite network of devices and places provides the starting situation. We then transform the network by measuring structural equivalence and creating a block model. Finally, it can be laid out spatially on top of a map (e.g. a street map or a floor plan of a building).

We have demonstrated this method in chapter 10 with three scenarios: the PerCom06 conference in Pisa, the CeBIT computer fair 06 in Hanover (both with Bluetooth localization) and the daily routine of our proband in Bremen (using Cell-ID for localization).

Detection of Episodes In chapter 11, we introduced a different method to detect changes in social context on a short-term, which can be used to separate different episodes during the day. Our method is based on the differentiation of familiar (regularly detected) and strange (not repeatedly discovered) devices. By discovering dynamic in each of these classes of devices in proximity, we can distinguish situations of changing social situations from stable situations, separately for unknown and familiar people around us.

We have demonstrated the functioning of this method with a six-day dataset recorded at the Ubicomp05 conference, including conference days, workshop days and recreational time.

Characterization of Bluetooth Inquiries As a basis for the above methods, we have conducted experiments to characterize the properties of Bluetooth device inquiries. We conducted augmented gatecounts in different locations to measure and compare the degree to which we can detect people using Bluetooth. Chapter 6 contains controlled lab experiments and chapter 7 the gatecount experiments.

Dataset of Egocentric Proximity Data We collected a dataset of Bluetooth device inquiries with interesting features for our purposes. In contrast to the RealityMining dataset, we needed data from only one proband in a set of different contexts. The set contains seven different settings, recorded in a variety of locations around the world. It also contains gaps in time, so that we could show the effect of long-term developments.

Two subsets of data were augmented by a set of additional stationary scanners with known locations, which afford an allocentric perspective on the social situation. Another instance contains Cell-ID information in addition to the Bluetooth discoveries.

Soft- and Hardware for Measurement and Data Aggregation To realize the practical collection of the dataset and to conduct measurement of Bluetooth performance including the gatecounts, we developed a system of software tools.

A program for mobile phones was used to carry out the main data collection from the egocentric perspective. It provides a user interface to visualize and explore the current social context spontaneously. A number of embedded boards were assembled to create small stationary devices, which could be easily deployed at various settings. They provide greater control over the parameters of the Bluetooth protocol than the constrained API of mobile phones. The stationary devices also acted as data collectors for the mobile phones. Thus, the mobile phone program automatically uploads its data to the stationary devices, which again forwarded all data to a single server, where it was aggregated in a database. A special program on the server transforms and exports the data for the various analyses we carried out.

Chapter 5 outlines the details of the software and hardware we developed. The software is available under the GPL license to support further research. It can be downloaded at
`http://sourceforge.net/projects/wirelessrope` [202].

12.2. Future Work

Many questions come to mind, with the work presented here. Social context is a topic not extensively researched in computer science, yet. We have shown a number of illustrative examples with the goal to demonstrate the potential of the concept of social context. It is probably the interdisciplinary nature of social context, what makes this research difficult. As computer scientists, we often arrived at the limits of our technical perspective—and were tempted to cross the boundary. We think, that there is great potential in broad interdisciplinary research into this topic.

When it comes to the technical basis of this work, many developments have happened from when we started. The Bluetooth specification has received major updates, however the device inquiry procedure has not been in the focus. As we have argued in chapter 2, the device inquiry is more than a tedious necessity for connecting devices. With several improvements, it could be a convenient basis for proximity applications in general. To improve Bluetooth's applicability here, faster detection of other devices, precise distance measurements and privacy control would be needed.

In the meanwhile, a few years have passed since we conducted the augmented gatecount studies. It would be interesting to survey more locations and see how the Bluetooth penetration changed over time. Recently, the measurement of people's density has become of commercial interest, with data based on mobile phone localizations via WiFi [177]. How does the Bluetooth method compare in terms of precision?

The two methods dealing with temporal data (chapters 9 and 11) are relevant in supporting human episodic memory and automatic diary applications. Could we improve the Forget-Me-Not device this way?

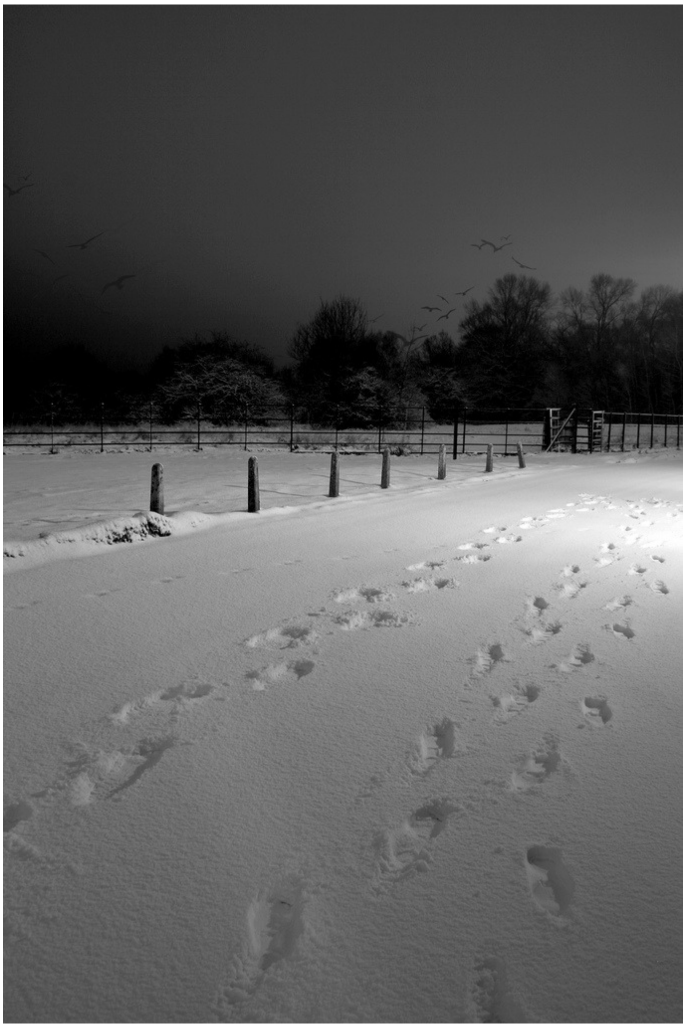
Although we have discussed how our methods relate to other approaches, e.g. the eigenbehavior method, we have not combined multiple approaches. Each method is able to reveal another aspect in the structure of our social fabric. Combinations should prove to be more powerful.

There is also potential to elaborate the methods presented here by using them on different data sets. The spatial mapping we have done in chapter 10 has only been demonstrated with sparse data (Routine05) or small areas (PerCom06). We would like to run it on larger datasets of combined Bluetooth and GPS data. It should be possible to refine the method to create areas with crisp geographical boundaries compared to our fuzzy Cell-ID areas. The large Cell-ID datasets from mobile operators would also be worth an analysis. Such data could create a social map of a large area, e.g. a city.

We believe, that such social maps will become an indispensable complement to street maps. In the last years, we have experienced a shift towards purely geographical navigation, manifested in the pervasive car navigation systems. It is time for the social navigation approach to be revived and to augment our geographic street maps. If there are many ways we could take through the city, which would we *want* to take? There might be neighborhoods we want to avoid. There might be areas where the probability to meet people of "our kind" is higher.

Let us follow the footprints in the snow.[9]

[9] More on social navigation in [73]



Photograph by miss_blackbutterfly[10]

[10] http://www.flickr.com/photos/blackbutterfly/3245462287/, accessed May 24th, 2010

Bibliography

[1] ACME Systems. Website. `http://www.acmesystems.it`. Accessed February 14th, 2008.

[2] Aka-Aki. Website. `http://www.aka-aki.com`. Accessed December 14th, 2007.

[3] Guiseppe Anastasi, Renata Bandelloni, Marco Conti, Franca Delmastro, Enrico Gregori, and Giovanni Mainetto. Experimenting an indoor Bluetooth-based positioning service. In *ICDCSW '03: Proceedings of the 23rd International Conference on Distributed Computing Systems*, pages 480–483, Los Alamitos, CA, USA, 2003. IEEE Computer Society.

[4] AOL Digital City. Website. `http://aol.digitalcity.com`. Accessed June 11th, 2007.

[5] Avalanche Transceiver. Website. `http://www.backcountryaccess.com/english/products/tracker_dts.php`. Accessed December 15th, 2007.

[6] Paramvir Bahl and Venkata N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of 19th IEEE Conference on Computer Communications (INFOCOM)*, pages 775–784, Washington, DC, USA, 2000. IEEE Computer Society.

[7] Magdalena Balazinska and Paul Castro. Characterizing mobility and network usage in a corporate wireless local-area network. In *MobiSys '03: Proceedings of the 1st international conference on Mobile Systems, Applications and Services*, pages 303–316, New York, NY, USA, 2003. ACM.

[8] Trevor J. Barnes. A paper related to everything but more related to local things. *Annals of the Association of American Geographers*, 94(2):278–283, June 2004.

[9] Russell Beale. Supporting social interaction with smart phones. *IEEE Pervasive Computing*, 4(2):35–41, 2005.

[10] Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[11] Bluediving. Website. http://bluediving.sourceforge.net. Accessed May 22nd, 2010.

[12] Bluetooth SIG. Bluetooth assigned numbers. http://www.bluetooth.org/assigned-numbers/baseband.htm. Accessed Dec. 13th, 2007.

[13] Bluetooth SIG. Specification of the Bluetooth system, November 2004.

[14] Bluetooth SIG. Java APIs for Bluetooth wireless technology, September 2005.

[15] R. Darrell Bock and Suraya Zahid Husain. Factors of the tele: A preliminary report. *Sociometry*, 15(3/4):206–219, 1952.

[16] Steve Borgatti, Martin Everett, and Linton C. Freeman. UCINET for Windows: Software for social network analysis. Harvard, MA: Analytic Technologies, 2002.

[17] Richard Borovoy, Fred Martin, Mitchel Resnick, and Brian Silverman. GroupWear: Nametags that tell about relationships. In *CHI '98: Conference Summary on Human Factors in Computing Systems*, pages 329–330, New York, NY, USA, 1998. ACM.

[18] Richard Borovoy, Fred Martin, Sunil Vemuri, Mitchel Resnick, Brian Silverman, and Chris Hancock. Meme tags and community mirrors: Moving from conferences to collaboration. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 159–168, New York, NY, USA, 1998. ACM.

[19] Bremen online. Website. http://www.bremen.de. Accessed June 11th, 2007.

[20] P. J. Brown, J. D. Bovey, and X. Chen. Context-aware applications: from the laboratory to the marketplace. *IEEE Personal Communications*, 4(5):58–64, October 1997.

[21] Ronald S. Burt. Positions in networks. *Social Forces*, 55(1):93–122, September 1976.

[22] Tracy Camp, Jeff Boleng, and Vanessa Davies. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing*, 2(5):483–502, August 2002.

[23] Andrew Campbell. Mobile phone sensing is the next big thing! Keynote at ACM MobiOpp 2010. http://www.cs.dartmouth.edu/~campbell/mobiopp-2010.pdf, Feb. 2010. Accessed May 24th, 2010.

[24] Andrew Campbell, Shane Eisenman, Nicholas Lane, Emiliano Miluzzo, and Ronald Peterson. People-centric urban sensing. In *WICON '06: Proceedings of the 2nd Annual International Workshop on Wireless Internet*, New York, NY, USA, 2006. ACM.

[25] Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, and James Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proceedings of 25th IEEE Conference on Computer Communications (INFOCOM)*, Washington, DC, USA, 2006. IEEE Computer Society.

[26] Charmed Technology. Website. `http://www.charmed.com/products/charmbadge.html`. Accessed December 14th, 2007.

[27] Guanling Chen and David Kotz. A survey of context-aware mobile computing research. Technical report, Dartmouth College, Hanover, NH, USA, 2000.

[28] Mike Chen, Dirk Haehnel, Jeffrey Hightower, Timothy Sohn, Anthony LaMarca, Ian Smith, Dmitri Chmelev, Jeff Hughes, and Fred Potter. Practical metropolitan-scale positioning for GSM phones. In Paul Dourish and Adrian Friday, editors, *Proceedings of the 8th International Conference on Ubiquitous Computing (Ubicomp 2006)*, pages 225–242, Berlin, Heidelberg, 2006. Springer-Verlag.

[29] Tanzeem K. Choudhury. *Sensing and Modeling Human Networks*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, February 2004.

[30] Cityware Facebook App. Website. `http://apps.facebook.com/cityware`. Accessed December 15th, 2007.

[31] Cityware Project. Website. `http://www.cityware.org.uk/`. Accessed April 30th, 2010.

[32] Come Out and Play Festival. Website. `http://www.comeoutandplay.org`. Accessed December 12th, 2007.

[33] Donna Cox, Volodymyr Kindratenko, and David Pointer. IntelliBadge: Towards providing location-aware value-added services at academic conferences. In *Proceedings of the 5th International Conference on Ubiquitous Computing (Ubicomp 2003)*, pages 264–280, Berlin, Heidelberg, 2003. Springer-Verlag.

[34] Crawdad Project. Website. `http://crawdad.cs.dartmouth.edu`. Accessed December 13th, 2007.

[35] Michael R. Curry, David J. Phillips, and Priscilla M. Regan. Emergency response systems and the creeping legibility of people and places. *The Information Society*, 20(5):357–369, November 2004.

[36] James A. Davis, Andrew H. Fagg, and Brian N. Levine. Wearable computers as packet transport mechanisms in highly partitioned ad-hoc networks. In *ISWC '01: Proceedings of the 5th IEEE International Symposium on Wearable Computers*, pages 141–148, Los Alamitos, CA, USA, 2001. IEEE Computer Society.

[37] Janet Davis, Peyina Lin, Alan Borning, Bataya Friedman, Peter H. Kahn Jr., and Paul A. Waddell. Simulations for urban planning: Designing for human values. *IEEE Computer*, 39(9):66–72, September 2006.

[38] Department of Economic and Social Affairs of the United Nations Secretariat. World urbanization prospects: The 2007 revision. `http://esa.un.org/unup`, June 2008. Accessed June 20th, 2010.

[39] Anind K. Dey. Context-aware computing: The CyberDesk project. In *Proceedings of AAAI '98 Spring Symposium on Intelligent Environments (AAAI '98 IE)*, pages 51–54, Menlo Park, CA, USA, 1998. AAAI Press.

[40] Anind K. Dey, Gregory Abowd, and Daniel Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Comupter Interaction*, 16(2):97–166, 2001.

[41] Dodgeball. Website. `http://www.dodgeball.com`. Accessed December 15th, 2007.

[42] Paul Dourish. What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8(1):19–30, 2004.

[43] Émile Durkheim. *The rules of the sociological method*. The Free Press, New York, NY, USA, 1982.

[44] Nathan Eagle. *Machine Perception and Learning of Complex Social Systems*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, May 2005.

[45] Nathan Eagle and Alex Pentland. Social serendipity: Mobilizing social software. *IEEE Pervasive Computing*, 4(2):28–34, 2005.

[46] Nathan Eagle and Alex Pentland. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, November 2006.

[47] EKAHAU. Website. `http://www.ekahau.com`. Accessed December 21st, 2007.

[48] Ian Elcoate, Jim Longstaff, and Paul Massey. Location privacy in multiple social contexts. In *Workshop on Privacy, Trust and Identity Issues for Ambient Intelligence at Pervasive*, 2006.

[49] Scott Elrod, Gene Hall, Rick Costanza, Michael Dixon, and Jim Des Rivieres. Responsive office environments. *Communications of the ACM*, 36(7):84–85, 1993.

[50] Enhanced 911 - Wireless Services. Website. `http://www.fcc.gov/pshs/911/enhanced911`. Accessed December 15th, 2007.

[51] Matthias Esbjörnsson, Oskar Juhlin, and Matthias Östergren. Traffic encounters and Hocman: Associating motorcycle ethnography with design. *Personal and Ubiquitous Computing*, 8:92–99, May 2004.

[52] Claude S. Fischer. *The Urban Experience*. Harcourt Brace Jovanovich, San Diego, CA, USA, 1976.

[53] Marcus Foth. Facilitating social networking in inner-city neighborhoods. *IEEE Computer*, 39(9):44–50, September 2006.

[54] Linton C. Freeman. *Models and Methods in Social Network Analysis*, chapter Graphic Techniques for Exploring Social Network Data, pages 248–269. Cambridge University Press, Cambridge, UK, 2005.

[55] Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Software-Practice and Experience*, 21(11):1129–1164, 1991.

[56] Werner Fuchs, Rolf Klima, Rüdiger Lautmann, Otthein Rammstedt, and Hanns Wienold, editors. *Lexikon zur Soziologie*. Westdeutscher Verlag, Leverkusen, Germany, 2nd edition, 1978.

[57] Jonathan Gips and Alex Pentland. Mapping human networks. In *PERCOM '06: Proceedings of the 4th Annual IEEE International Conference on Pervasive Computing and Communications*, pages 159–168, Washington, DC, USA, 2006. IEEE Computer Society.

[58] Google. Google maps with my location (beta). `http://www.google.com/gmm/mylocation.html`. accessed, December 20th, 2007.

[59] Stephen Graham. *Technocities*, chapter Towards Urban Cyberspace Planning: Grounding the Global through Urban Telematics Policy and Planning, pages 9–33. Sage Publications, Thousand Oaks, CA, USA, 1999.

[60] Gary Gumpert and Susan Drucker. Privacy, predictability or serendipity and digital cities. In Makoto Tanabe, Peter van den Besselaar, and Toru Ishida, editors, *Digital Cities II: Computational and Sociological Approaches*, pages 26–40, Berlin, Heidelberg, 2002. Springer-Verlag.

[61] Haggle Project. Website. `http://www.haggleproject.org`. Accessed December 13th, 2007.

[62] Edward T. Hall. *The Hidden Dimension*. Doubleday, New York, NY, USA, 1966.

[63] Keith N. Hampton. Grieving for a lost network collective action in a wired suburb. *The Information Society*, 19(5):417–428, 2003.

[64] Hartmut Häussermann and Walter Siebel. *Stadtsoziologie*. Campus Verlag, Frankfurth am Main, Germany, 2004.

[65] Michael Hebbert. The street as locus of collective memory. *Environment and Planning D: Society and Space*, 23(4):581–596, 2005.

[66] Jeffrey Hightower and Gaetano Borriello. Location systems for ubiquitous computing. *IEEE Computer*, 34(8):57–66, August 2001.

[67] Jeffrey Hightower, Sunny Consolvo, Anthony LaMarca, Ian Smith, and Jeff Hughes. Learning and recognizing the places we go. In *Proceedings of the 7th International Conference on Ubiquitous Computing (Ubicomp 2005)*, pages 105–122, Berlin, Heidelberg, 2005. Springer-Verlag.

[68] Jeffrey Hightower, Anthony LaMarca, and Ian Smith. Practical lessons from Place Lab. *IEEE Pervasive Computing*, 5:32–39, 2006.

[69] Jeffrey Hightower, Chris Vakili, Gaetano Borriello, and Roy Want. Design and calibration of the SpotON ad-hoc location sensing system. Technical report, University of Washington, Seattle, WA, USA, August 2001.

[70] Bill Hillier, Richard Burdett, John Peponis, and Alan Penn. Creating life: Or, does architecture determine anything? *Architecture & Behaviour*, 3(3):233–250, 1987.

[71] Bill Hillier and Julienne Hanson. *The Social Logic of Space*. Cambridge University Press, Cambridge, UK, 1984.

[72] Lars Erik Holmquist, Jennica Falk, and Joakim Wigström. Supporting group collaboration with interpersonal awareness devices. *Personal Technologies*, 3(1–2):13–21, 1999.

[73] Kristina Höök, David Benyon, and Alan J. Munro, editors. *Designing Information Spaces: The Social Navigation Approach*. Springer-Verlag, Berlin, Heidelberg, 2003.

[74] Albert Huang and Larry Rudolph. A privacy conscious Bluetooth infrastructure for location aware computing. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA, January 2005.

[75] Elaine M. Huang, Michael Terry, Elizabeth D. Mynatt, Kent Lyons, and Alan Chen. Distributing event information by simulating word-of-mouth exchanges. In *Mobile HCI '02: Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction*, pages 60–68, Berlin, Heidelberg, 2002. Springer-Verlag.

[76] R. Robert Huckfeldt. Social contexts, social networks, and urban neighborhoods: Environmental constraints on friendship choice. *American Journal of Sociology*, 89(3):651–669, November 1983.

[77] Pan Hui, Augustin Chaintreau, James Scott, Richard Gass, Jon Crowcroft, and Christophe Diot. Pocket switched networks and human mobility in conference environments. In *Proceeding of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN '05)*, pages 244–251, New York, NY, USA, 2005. ACM.

[78] Chris Hurley, Michael Puchol, Russ Rogers, and Frank Thornton. *WarDriving: Drive, Detect, Defend, a Guide to Wireless Security*. Syngress Publishing, Burlington, MA, USA, 2004.

[79] Toru Ishida. Digital city Kyoto. *Communications of the ACM*, 45(7):76–81, 2002.

[80] Toru Ishida, Jun ichi Akahani, Kaoru Hiramatsu, Katherine Isbister, Stefan Lisowski, Hideyuki Nakanishi, Masayuki Okamoto, Yasuhiko Miyazaki, and Ken Tsutsuguchi. Digital city Kyoto: Towards a social information infrastructure. In Matthias Klusch, Onn M. Shehory, and Gerhard Weiß, editors, *CIA '99: Proceedings of the 3rd International Workshop on Cooperative Information Agents III*, pages 34–46, Berlin, Heidelberg, 1999. Springer-Verlag.

[81] Yukari Iwatani. Love: Japanese style. Wired: `http://www.wired.com/news/culture/0,1284,12899,00.html`, 1998. Accessed: November 24th, 2006.

[82] Margot Jacobs, Lalya Gaye, and Lars Erik Holmquist. Tejp: Ubiquitous computing as expressive means of personalising public space. In *Adjunct Proceedings of Ubicomp '03*, 2003.

[83] Jaiku. Website. `http://www.jaiku.com`. Accessed December 15th, 2007.

[84] Java ME. Website. `http://www.oracle.com/technetwork/java/javame`. Accessed Sept. 10th, 2010.

[85] Tomihisa Kamada and Satoru Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31:7–15, April 1989.

[86] Marije Kanis, Niall Winters, Stefan Agamanolis, Anna Gavin, and Cian Cullinan. Toward wearable social networking with iBand. In *CHI '05: Extended Abstracts on Human Factors in Computing Systems*, pages 1521–1524, New York, NY, USA, 2005. ACM.

[87] Nicky Kern, Bernt Schiele, Holger Junker, Paul Lukowicz, and Gerhard Tröster. Wearable sensing to annotate meeting recordings. *Personal and Ubiquitous Computing*, 7(5):263–274, September 2003.

[88] Ruth Kikin-Gil. Affective is effective: how information appliances can mediate relationships within communities and increase one's social effectiveness. *Personal and Ubiquitous Computing*, 10(2):77–83, 2006.

[89] Tim Kindberg and Timothy Jones. "merolyn the phone": A study of Bluetooth naming practices. In *Proceedings of the 9th International Conference on Ubiquitous Computing (Ubicomp 2007)*, pages 318–355, Berlin, Heidelberg, 2007. Springer-Verlag.

[90] Jesper Kjeldskov and Jeni Paay. Public pervasive computing: Making the invisible visible. *IEEE Computer*, 39(9):60–65, September 2006.

[91] Anders Kofod-Petersen and Agnar Aamodt. Case-based situation assessment in a mobile context-aware system. In Antonio Krüger and Rainer Malaka, editors, *Artificial intelligence in Mobile Systems 2003 (AIMS)*, pages 41–49. Universität des Saarlandes, 2003.

[92] Gerd Kortuem and Zary Segall. Wearable communities: Augmenting social networks with wearable computers. *IEEE Pervasive Computing*, 2(1):71–78, January–March 2003.

[93] Vassilis Kostakos, Tom Nicolai, Eiko Yoneki, Eamonn O'Neill, Holger Kenn, and Jon Crowcroft. Understanding and measuring the urban pervasive infrastructure. *Personal and Ubiquitous Computing*, 5(13):355–364, 2009.

[94] Vassilis Kostakos, Eamon O'Neill, Alan Penn, George Roussos, and Dikaios Papadongonas. Brief encounters: Sensing, modeling and visualizing urban mobility and copresence networks. *ACM Trans. Comput.-Hum. Interact.*, 17(1):1–38, 2010.

[95] Vassilis Kostakos and Eamonn O'Neill. Quantifying the effects of space on encounter. In *Proceedings of the 6th International Space Syntax Symposium*, Istanbul, Turky, 2007. ITU Faculty of Architecture.

[96] Vassilis Kostakos, Eamonn O'Neill, and Alan Penn. Designing urban pervasive systems. *IEEE Computer*, 39(9):52–59, September 2006.

[97] Robert Kraut, Sara Kiesler, Bonka Boneva, Jonathan Cummings, Vicki Helgeson, and Anne Crawford. Internet paradox revisited. *Journal of Social Issues*, 58(1):49–74, 2002.

[98] Victor V. Kryssanov, Masayuki Okabe, Koh Kakusho, and Michihiko Minoh. Communication of social agents and the digital city – a semiotic pespective. In Makoto Tanabe, Peter van den Besselaar, and Toru Ishida, editors, *Digital Cities II: Computational and Sociological Approaches*, pages 56–70, Berlin, Heidelberg, 2002. Springer-Verlag.

[99] G. Lachapelle, O. Mezentsev, J. Collin, and G. MacGougan. Pedestrian and vehicular navigation under signal masking using integrated HSGPS and self contained sensor technologies. In *Proceedings of the 11th World Congress, International Association of Institutes of Navigation*. International Association of Institutes of Navigation, October 2003.

[100] Anthony LaMarca, Yatin Chawathe, Sunny Consolvo, Jeffrey Hightower, Ian Smith, James Scott, Timothy Sohn, James Howard, Jeff Hughes, Fred Potter, Jason Tabert, Pauline Powledge, Gaetano Borriello, and Bill Schilit. Place Lab: Device positioning using radio beacons in the wild. In *Proceedings of the 3rd International Conference on Pervasive Computing (Pervasive 2005)*, pages 116–133, Berlin, Heidelberg, 2005. Springer-Verlag.

[101] Michael G. Lamming and Denis Bohm. SPECs: Another approach to human context and activity sensing research, using tiny peer-to-peer wireless computers. In *Proceedings of the 5th International Conference on Ubiquitous Com-*

puting (Ubicomp 2003), pages 192–199, Berlin, Heidelberg, 2003. Springer-Verlag.

[102] Michael G. Lamming, Denis Bohm, and Robert N. Mayo. Lessons learned from a personal sensing architecture. Technical report, HP Laboratories, Palo Alto, CA, USA, August 2005.

[103] Michael G. Lamming and Mike Flynn. Forget-me-not: Intimate computing in support of human memory. In *Proceedings of the FRIEND21 Symposium on Next Generation Human Interfaces*, 1994.

[104] Michael G. Lamming and William M. Newman. Activity-based information retrieval: Technology in support of personal memory. In *Proceedings of the IFIP 12th World Computer Congress on Personal Computers and Intelligent Systems - Information Processing '92 - Volume 3*, pages 68–81, Amsterdam, Netherlands, 1992. North-Holland Publishing Co.

[105] Giles Lane. Urban tapestries: Wireless networking, public authoring and social knowledge. *Personal and Ubiquitous Computing*, 7:169–175, 2003.

[106] Angela Y. Lee. The mere exposure effect: An uncertainty reduction explanation revisited. *Personality and Social Psychology Bulletin*, 27(10):1255–1266, 2001.

[107] Todd Litman. London congestion pricing: Implications for other cities. Victoria Transport Policy Institute, `http://www.vtpi.org/london.pdf`, January 2006. Accessed June 13th, 2007.

[108] Benyuan Liu, Peter Brass, Olivier Dousse, Philippe Nain, and Don Towsley. Mobility improves coverage of sensor networks. In *MobiHoc '05: Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 300–308, New York, NY, USA, 2005. ACM.

[109] Loopt. Website. `https://www.loopt.com`. Accessed December 15th, 2007.

[110] Francois Lorrain and Harrison C. White. Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1:49–80, 1971.

[111] Anil Madhavapeddy and Alastair Tse. A study of Bluetooth propagation using accurate indoor location mapping. In *Proceedings of the 7th International Conference on Ubiquitous Computing (Ubicomp 2005)*, pages 105–122, Berlin, Heidelberg, 2005. Springer-Verlag.

[112] Mage Display Software. Website. `http://kinemage.biochem.duke.edu/software/mage.php`. Accessed Sept. 10th, 2010.

[113] Scott D. Mainwaring, Ken Anderson, and Michele Chang. Living for the global city: Mobile kits, urban interfaces, and ubicomp. In *Proceedings of the 7th International Conference on Ubiquitous Computing (Ubicomp 2005)*, pages 269–286, Berlin, Heidelberg, 2005. Springer-Verlag.

[114] Steve Mann. *Intelligent Image Processing*. John Wiley and Sons, Hoboken, NJ, USA, 2001.

[115] Steve Mann. Continuous lifelong capture of personal experience with EyeTap. In *CARPE '04: Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 1–21, New York, NY, USA, 2004. ACM.

[116] Steve Mann. "Sousveillance": inverse surveillance in multimedia imaging. In *MULTIMEDIA '04: Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 620–627, New York, NY, USA, 2004. ACM.

[117] Giuseppe Mantovani. Social context in HCI: A new framework for mental models, cooperation, and communication. *Cognitive Science*, 20:237–269, 1996.

[118] Michael Massimi, Craig H. Ganoe, and John M. Carroll. Scavenger hunt: An empirical method for mobile collaborative problem-solving. *IEEE Pervasive Computing*, 6:81–87, 2007.

[119] Joseph F. McCarthy. Active environments: Sensing and responding to groups of people. *Personal and Ubiquitous Computing*, 5(1):75–77, 2001.

[120] Joseph F. McCarthy, David H. Nguyen, Al Mamunur Rashid, and Suzanne Soroczak. Proactive displays & the experience UbiComp project. *ACM SIGGROUP Bulletin*, 23(3):38–41, 2003.

[121] Marvin McNett and Geoffrey M. Voelker. Access and mobility of wireless PDA users. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(2):40–55, 2005.

[122] Metro Sense Project. Website. `http://metrosense.cs.dartmouth.edu`. Accessed April 30th, 2010.

[123] Stanley Milgram. *The Individual in a Social World: Essays and Experiments*. Addison-Wesley, 1977.

[124] Mobile Environmental Sensing System Across a Grid Environment (MESSAGE) Project. Website. http://bioinf.ncl.ac.uk/message/. Accessed April 30th, 2010.

[125] Mobiluck. Website. http://www.mobiluck.com. Accessed December 14th, 2007.

[126] Marko Modsching, Ronny Kramer, and Klaus ten Hagen. Field trial on GPS accuracy in a medium size city: The influence of built-up. In *Proceedings of the 3rd Workshop on Positioning, Navigation and Communication (WPNC'06)*, pages 209–218, Aachen, Germany, 2006. Shaker Verlag.

[127] Mologogo. Website. http://www.mologogo.com. Accessed December 15th, 2007.

[128] Jakob L. Moreno. Sociometry in relation to other social sciences. *Sociometry*, 13(1):63–75, February 1950.

[129] Moviemol Molecular Animation Program. Website. http://www.ifm.liu.se/compchem/moviemol/moviemol.html. Accessed Sept. 10th, 2010.

[130] Gina Neff. The changing place of cultural production: The location of social networks in a digital media industry. *The Annals of the American Academy of Political and Social Science*, 597(1):134–152, 2005.

[131] New York Times. A peek into Netflix queues. http://www.nytimes.com/interactive/2010/01/10/nyregion/20100110-netflix-map.html, January 2010. Accessed May 10th, 2010.

[132] Lionel M. Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P. Patil. LANDMARC: Indoor location sensing using active RFID. *Wireless Networks*, 10:701–710, 2004.

[133] Tom Nicolai, Nils Behrens, and Heidi Thielemann. "Be a Freeporter": Enabling a mobile news publishing community. In Karenza Moore and Jason Rutter, editors, *Mobile Entertainment: User-Centred Perspectives*, pages 254–272, Manchester, UK, 2004. CRIC.

[134] Tom Nicolai and Holger Kenn. Towards detecting social situations with Bluetooth. In *Adjunct Proceedings of Ubicomp '06*, 2006.

[135] Tom Nicolai and Holger Kenn. About the relationship between people and discoverable Bluetooth devices in urban environments. In Peter Han Joo Chong and Adrian David Cheok, editors, *Mobility '07: Proceedings of the 4th International Conference on Mobile Technology, Applications and Systems*, pages 72–78, New York, NY, USA, 2007. ACM.

[136] Tom Nicolai, Eiko Yoneki, Nils Behrens, and Holger Kenn. Exploring social context with the Wireless Rope. In Robert Meersman, Zahir Tari, and Pilar Herrero, editors, *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Part I*, pages 874–883, Berlin, Heidelberg, 2006. Springer-Verlag.

[137] Norman Nie. Sociability, interpersonal relations, and the Internet: Reconciling conflicting findings. *American Behavioral Scientist*, 45(3):420–435, 2001.

[138] Norman Nie and Lutz Erbring. *The digital divide: facing a crisis or creating a myth?*, chapter Internet and Society: A Preliminary Report, pages 269–271. MIT Press, Cambridge, MA, USA, 2001.

[139] Nike+. Website. `http://www.nikeplus.com`. Accessed December 12th, 2007.

[140] Kazushi Nishimoto, Tadao Maekawa, Yukio Tada, Kenji Mase, and Ryohei Nakatsu. Networked wearable musical instruments will bring a new musical culture. In *ISWC '01: Proceedings of the 5th IEEE International Symposium on Wearable Computers*, pages 55–62, Los Alamitos, CA, USA, 2001. IEEE Computer Society.

[141] Nokia Sensor. Website. `http://web.nokia.de/de/service/software/sensor/startseite/168744.html`. Accessed December 14th, 2007.

[142] Wouter De Nooy, Andrej Mrvar, and Vladimir Batagelj. *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, Cambridge, UK, 2005.

[143] nTAG Interactive. Website. `http://www.ntag.com`. Accessed December 14th, 2007.

[144] Sarah Olofsson, Veronica Carlsson, and Jessica Sjölander. The friend locator: Supporting visitors at large-scale events. *Personal and Ubiquitous Computing*, 10(2–3):84–89, April 2006.

[145] Eamonn O'Neill, Vassilis Kostakos, Tim Kindberg, Ava Fatah gen. Schieck, Alan Penn, Danae Stanton Fraser, and Tim Jones. Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In Paul Dourish and Adrian Friday, editors, *Proceedings of the 8th International Conference on Ubiquitous Computing (Ubicomp 2006)*, pages 315–332, Berlin, Heidelberg, 2006. Springer-Verlag.

[146] Leysia Palen and Paul Dourish. Unpacking "privacy" for a networked world. *CHI Letters*, 5(1):129–136, 2003.

[147] Dikaios Papadogkonas, George Roussos, and Mark Levene. Ranking significant ubiquitous computing trails. In *2nd International Workshop on Exploiting Context Histories in Smart Environments (EICHSE) at UbiComp '06, Online Proceedings*, 2006.

[148] Shwetak N. Patel, Julie A. Kientz, Gillian R. Hayes, Sooraj Bhat, and Gregory Abowd. Farther than you may think: An empirical investigation of the proximity of users to their mobile phones. In Paul Dourish and Adrian Friday, editors, *Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp 2006)*, pages 123–140, Berlin, Heidelberg, 2006. Springer-Verlag.

[149] Eric Paulos and Elizabeth Goodman. The familiar stranger: Anxiety, comfort, and play in public places. In *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 223–230, New York, NY, USA, 2004. ACM.

[150] Alex Pentland. Smart rooms. *Scientific American*, 274(4):68–76, April 1996.

[151] Place Lab. Website. `http://www.placelab.org`. Accessed December 18th, 2007.

[152] Plazes. Website. `http://plazes.com`. Accessed December 15th, 2007.

[153] Juval Portugali. Self-organizing cities. *Futures*, 29(4/5):353–380, 1997.

[154] Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. The Cricket location-support system. In *Proceedings of the 6th Annual ACM International Conference on Mobile Computing and Networking (MOBICOM)*, pages 32–43, New York, NY, USA, 2000. ACM.

[155] Qiro. Website. `https://www.myqiro.de`. Accessed December 15th, 2007.

[156] Mika Raento, Antti Oulasvirta, Renaud Petit, and Hannu Toivonen. ContextPhone: A prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing*, 4(2):51–59, 2005.

[157] Reality Mining Project. Website. `http://reality.media.mit.edu`. Accessed December 13th, 2007.

[158] Howard Rheingold. *Smart Mobs: The Next Social Revolution*. Perseus Publishing, 2002.

[159] Bradley J. Rhodes. The wearable remembrance agent: A system for augmented memory. In *ISWC '97: Proceedings of the 1st IEEE International Symposium on Wearable Computers*, pages 123–128, Washington, DC, USA, 1997. IEEE Computer Society.

[160] Oriana Riva and Cristian Borcea. The Urbanet revolution: Sensor power to the people! *IEEE Pervasive Computing*, 6(2):41–49, April 2007.

[161] Nick Ryan, Jason Pascoe, and David Morse. Enhanced reality fieldwork: the context-aware archaeological assistant. In V. Gaffney, M. van Leusen, and S. Exxon, editors, *Computer Applications in Archaeology (1997)*, British Archaeological Reports, Oxford, UK, 1997. Tempus Reparatum.

[162] Nitin Sawhney, Sean Wheeler, and Chris Schmandt. Aware community portals: Shared information appliances for transitional spaces. *Personal and Ubiquitous Computing*, 5(1):66–70, 2001.

[163] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *WMCSA '94: Proceedings of the 1994 1st Workshop on Mobile Computing Systems and Applications*, pages 85–90, Washington, DC, USA, 1994. IEEE Computer Society.

[164] Bill Schilit and Marvin Theimer. Disseminating active map information to mobile hosts. *IEEE Network*, 8(5):22–32, 1994.

[165] Albrecht Schmidt. *Ubiquitous Computing – Computing in Context*. PhD thesis, Computing Department, Lancaster University, UK, 2002.

[166] Albrecht Schmidt, Kofi Asante Aidoo, Antti Takaluoma, Urpo Tuomela, Kristof Van Laerhoven, and Walter Van de Velde. Advanced interaction in context. In H.-W. Gellersen, editor, *HUC '99*, LNCS, pages 89–101, Berlin Heidelberg, 1999. Springer Verlag.

[167] Jay Schneider, Gerd Kortuem, Joe Jager, Steve Fickas, and Zary Segall. Disseminating trust information in wearable communities. *Personal and Ubiquitous Computing*, 4(4):245–248, 2000.

[168] Andrew Sears, Min Lin, Julie Jacko, and Yan Xiao. When computers fade pervasive computing and situationally-induced impairments and disabilities. In Julie Jacko and Constantine Stephanidis, editors, *Human-Computer Interaction: Theory and Practice (Part II)*, pages 1298–1302, Mahwah, New Jersey, USA, 2003. Lawrence Erlbaum Associates.

[169] Christian Seitz, Michael Berger, and Bernhard Bauer. Towards a general approach to mobile profile based distributed grouping. *Personal and Ubiquitous Computing*, 9(2):90–99, 2005.

[170] Sensor Planet. Website. `http://www.sensorplanet.org`. Accessed June 8th, 2007.

[171] Jeremy Shaffer, Daniel P. Siewiorek, and Asim Smailagic. Analysis of movement and mobility of wireless network users. In *ISWC '05: Proceedings of the 9th IEEE International Symposium on Wearable Computers*, pages 60–67, Washington, DC, USA, 2005. IEEE Computer Society.

[172] Irinia Shklovski and Michele Chang. Urban computing – navigating space and context. *IEEE Computer*, 39(9):36–37, September 2006.

[173] Frank Siegemund and Michael Rohs. Rendezvous layer protocols for Bluetooth-enabled smart devices. *Personal and Ubiquitous Computing*, 7(2):91–101, 2003.

[174] Bradley A. Singletary and Thad Starner. Symbiotic interfaces for wearable face recognition. In *HCII2001 Workshop On Wearable Computing*, pages 141–148, New Orleans, LA, August 2001.

[175] Sinus Sociovision. Informationen zu den Sinus-Milieus 2007. `http://www.sinus-sociovision.de/Download/informationen012007.pdf`, 2007. Accessed June 13th, 2007.

[176] Skyhook Wireless. Website. `http://www.skyhookwireless.com`. Accessed May 2nd, 2010.

[177] Skyhook Wireless: Spotrank. Website. `http://www.skyhookwireless.com/spotrank`. Accessed May 24th, 2010.

[178] Dan Smith, Ling Ma, and Nick Ryan. Acoustic environment as an indicator of social and physical context. *Personal and Ubiquitous Computing*, 10(4):241–254, 2006.

[179] Ian Smith, Sunny Consolvo, Anthony LaMarca, Jeffrey Hightower, James Scott, Timothy Sohn, Jeff Hughes, Giovanni Iachello, and Gregory Abowd. Social disclosure of place: From location technology to communication practices. In *Proceedings of the 3rd International Conference on Pervasive Computing (Pervasive 2005)*, pages 134–151, Berlin, Heidelberg, 2005. Springer-Verlag.

[180] Marc A. Smith. Some social implications of ubiquitous wireless networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 4(2):25–36, 2000.

[181] Thad Starner. The challenges of wearable computing: Part 1. *IEEE Micro*, 21(4):44–52, July/August 2001.

[182] Symbian Foundation Community Home. Website. `http://www.symbian.org`. Accessed Nov. 15th, 2010.

[183] Sakari Tamminen, Antti Oulasvirta, Kalle Toiskallio, and Anu Kankainen. Understanding mobile contexts. *Personal and Ubiquitous Computing*, 8(2):135–143, 2004.

[184] Michael Terry, Elizabeth D. Mynatt, Kathy Ryall, and Darren Leigh. Social net: Using patterns of physical proximity over time to infer shared interests. In *CHI '02: Extended Abstracts on Human Factors in Computing Systems*, pages 816–817, New York, NY, USA, 2002. ACM.

[185] Waldo R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240, 1970. Supplement: Proceedings.

[186] Andre Torre and Alain Rallet. Proximity and localization. *Regional Studies*, 39(1):47–59, January 2005.

[187] Alasdair Turner and Alan Penn. Encoding natural movement as an agent-based system: An investigation into human pedestrian behaviour in the built environment. *Environment and Planning B: Planning and Design*, 29(4):473–490, 2002.

[188] Ubisense. Website. `http://www.ubisense.net`. Accessed December 22nd, 2007.

[189] Urban Sensing Project. Website. http://research.cens.ucla.edu/projects/2006/systems/Urban_Sensing. Accessed December 12th, 2007.

[190] Peter van den Besselaar and Dennis Beckers. Demographics and sociographics of the digital city. In Toru Ishida, editor, *Community Computing and Support Systems, Social Interaction in Networked Communities*, pages 108–124, Berlin, Heidelberg, 1998. Springer-Verlag.

[191] Roy Want, Andy Hopper, Veronica Falcao, and Jonathan Gibbons. The active badge location system. *ACM Transactions on Information Systems*, 10(1):91–102, January 1992.

[192] Andy Ward, Alan Jones, and Andy Hopper. A new location technique for the active office. *IEEE Personal Communications*, 4(5):42–47, October 1997.

[193] Stanley Wassermann and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK, 1994.

[194] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of “small world” networks. *Nature*, 393:440–442, June 1998.

[195] Alexandra Weilenmann and Lars Erik Holmquist. Hummingbirds go skiing: Using wearable computers to support social interaction. In *ISWC ’99: Proceedings of the 3rd IEEE International Symposium on Wearable Computers*, pages 191–192, Washington, DC, USA, 1999. IEEE Computer Society.

[196] Mark Weiser. The computer for the 21st century. *Scientific American*, 265:94–104, 1991.

[197] Mark Weiser. Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7):75–84, July 1993.

[198] Mark Weiser and John Seely Brown. The coming age of calm technology. Technical report, Xerox PARC, Palo Alto, CA, USA, October 1996.

[199] Barry Wellman. Little boxes, glocalization, and networked individualism. In Makoto Tanabe, Peter van den Besselaar, and Toru Ishida, editors, *Digital Cities II: Computational and Sociological Approaches*, pages 10–25, Berlin, Heidelberg, 2002. Springer-Verlag.

[200] WiFi Maps. Website. http://www.wifimaps.com. Accessed May 22nd, 2010.

[201] Amanda Williams and Paul Dourish. Imagining the city: The cultural dimensions of urban computing. *IEEE Computer*, 39(9):38–43, September 2006.

[202] WirelessRope. Website. `http://sourceforge.net/projects/wirelessrope`. Accessed Sept. 10th, 2010.

[203] Wowgadget.tv. Website. `http://www.wowgadget.tv`. Accessed Sept. 10th, 2010.

[204] Xin Xing, Tobias Warden, Tom Nicolai, and Otthein Herzog. SMIZE: A spontaneous ride-sharing system for individual urban transit. In *MATES'09: Proceedings of the 7th German Conference on Multiagent System Technologies*, pages 165–176, Berlin, Heidelberg, 2009. Springer-Verlag.

[205] Guanhua Yan and Stephan Eidenbenz. Modeling propagation dynamics of Bluetooth worms. In *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS '07)*, pages 42–51, Washington, DC, USA, 2007. IEEE Computer Society.

[206] Moustafa Youssef and Ashok K. Agrawala. The Horus WLAN location determination system. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys 2005)*, pages 205–218, New York, NY, USA, 2005. ACM.

[207] Robert B. Zajonc. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement*, 9(2):1–27, June 1968.

[208] ZigBee Standards Organization. ZigBee Specification, December 2006.